This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

Exploring individual differences in academic achievement Genetics, personality and the school environment

Smith-Woolley, Emily Dawn

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

Exploring individual differences in academic achievement: genetics, personality and the school environment

Emily Smith-Woolley

MRC Social, Genetic and Developmental Psychiatry Centre, King's College London, University of London

Submitted for the degree of Doctor of Philosophy in Behavioural Genetics

September 2018

ABSTRACT

Education is one of society's biggest investments. It aims to equip students with the skills and knowledge necessary to make their way in the world. Students' performance at school (educational achievement) can set them on very different life trajectories. Furthermore, differences in academic achievement are linked to variation in health, wellbeing, income and even mortality. Therefore, it is of societal importance to understand why individuals differ in academic achievement, and to explore their causes and correlates to improve education.

Differences in educational achievement are often attributed to the environment, for example the quality of a school a child attends. However, decades of behavioural genetic literature has shown that achievement is also substantially influenced by genetic factors. In order to increase understanding of the genetic and environmental influences on academic achievement, this thesis explores: the extent to which genetic variants associated with educational attainment explain differences in personality, and their relationship with educational achievement (Chapter 2); average DNA and achievement differences between students attending selective and non-selective schools (Chapter 3); the influence of school quality on educational achievement and student wellbeing (Chapter 4); and the genetic architecture of attainment and achievement beyond compulsory education into university (Chapter 5).

This thesis uses data from the Twins Early Development Study (TEDS). TEDS is a UK-representative sample of over 10,000 twin pairs followed longitudinally from age two to age 22, with a genotyped subsample of approximately 6,000 unrelated individuals. This thesis capitalises on both twin analysis and DNA-based methods to investigate the aetiology of achievement during secondary school and into university.

This thesis provides evidence that: 1) genetic effects of educational attainment relate to personality and motivation, and explain part of the covariance between personality and achievement; 2) genetic and achievement differences between students attending different school types are primarily due to the heritable characteristics involved in pupil admission, including general cognitive ability, socioeconomic status and prior achievement; 3) independently-rated school quality has little influence on educational achievement or student wellbeing during secondary school; and 4) genetic influences on achievement and attainment extend beyond compulsory education into university.

A discussion of these findings and their implications for teachers, policy-makers and parents is provided in the final chapter (Chapter 6), along with the conclusions that can be drawn from this body of work.

ACKNOWLEDGEMENTS

The research presented in this thesis uses data collected as part of the longitudinal Twins Early Development Study (TEDS). I gratefully acknowledge the ongoing contribution of the TEDS participants and their families. TEDS is supported by a program grant to Robert Plomin from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938) and the European Commission (602768; 295366).

I am thankful for the support and encouragement of my supervisor, Professor Robert Plomin. His open-door policy and willingness to chat through ideas, problems and plans, has been instrumental to the work in this thesis and to my positive attitude towards academia in general. In addition, I will be forever grateful to my second supervisor, Professor Yulia Kovas, whose enthusiasm, pragmatism and ability to inspire all those around her is the reason I pursued behavioural genetics in the first place. I would also like to thank three of my closest collaborators, Dr. Jean-Baptiste Pingault, Dr. Kathryn Asbury and Professor Philip Dale, whose guidance and statistical expertise has improved my research greatly.

TEDS would not be the brilliant study that it is without the dedication of the fantastic TEDS team – Rachel Ogden, Louise Webster and Andy McMillian – as well as the brilliant PhD students past and present. Not only are these individuals my colleagues, but also my friends, and a great source of support throughout the PhD. A special thank you to fellow research assistant, PhD student and dear friend, Ziada Ayorech, whose help, encouragement and support has been invaluable. I've been so lucky to share this PhD journey with such an inspirational person.

Many friends have helped in different ways over the past few years: from keeping me entertained (members of the Sweetcorn group), providing constant encouragement (Tiff), making work an absolute pleasure (Office gang) and always being there (Angels).

Finally, I would like to thank my family – my Mum and Dad, their partners, my sister, auntie and uncle and my Nanny for their continued support. Their unfaltering belief in me over the years has given me the courage to try anything. My last thank you goes to my husband, Joe. Thank you for the numerous walks around Tooting Common, for the cups of tea (that often went undrunk), for proofreading this thesis and for constantly making me laugh. You make my life one that is full of joy.

AUTHOR DECLARATION

Data used in this thesis are from the Twins Early Development Study. Data used in Chapters 2-4 were collected prior to my PhD. However, I helped collect the data used in Chapter 5. I was responsible for preparing the phenotypic and genetic data for the analyses and was responsible for all analyses conducted. To the best of my knowledge, the work presented here is original and my own work, except where acknowledged in the text.

Emily Smith-Woolley

INDEX

INDEX OF FIGURES

INDEX OF TABLES

Chapter 2 – Polygenic score for educational attainment captures DNA variants shared between personality-related traits and educational achievement

“The mind is not a vessel to be filled, but a fire to be kindled.”

- Plutarch

Chapter 1 – Introduction

BACKGROUND

By the time a student in the UK has reached the end of compulsory education at age 16, they will have received around 15,000 hours of schooling (Rutter, 1982) and it will have cost approximately £6,300 per year (Belfield, Crawford, & Sibieta, 2017). This makes education one of the most expensive environmental interventions routinely provided to children. Given this vast input, both in terms of time and expense, it is not surprising that society places high importance on individual academic achievement. Indeed, achievement in school can set students on very different life trajectories, with achievement differences associated with occupational status, income, health and even mortality (Deary & Johnson, 2010; Gathmann, Jürges, & Reinhold, 2015; Oreopoulos & Salvanes, 2011). Therefore, it is of societal importance to understand why individuals differ in academic achievement, and to explore causes and correlates, in order to improve education.

A wealth of literature points to genetics as one of the major systematic forces influencing individual differences in academic achievement. Twin studies have shown that additive genetic factors account for between 40-60% of the differences in academic achievement throughout development (Baker, Treloar, Reynolds, Heath, & Martin, 1996; Bartels, Rietveld, Van Baal, & Boomsma, 2002; Branigan, McCallum, & Freese, 2013; Kovas, Haworth, Dale, & Robert, 2007; Rimfeld et al., in press). Furthermore, recent genetic methods using DNA alone, can now explain up to 16% of the variance in academic achievement at the end of compulsory education (Allegrini et al., in press). This converging evidence supports substantial genetic influence on academic achievement. However, the same research also provides the best evidence for the importance of the environment. Only by accounting for genetics is it possible to accurately estimate the influence of the environment, in order to guide interventions, inform policy and improve education.

The work presented in this thesis aims to further understanding of academic achievement by: 1) exploring the contribution of genetics to the association between personality domains and achievement; 2) investigating the influence of school environments, such as school quality and school type, on academic achievement, wellbeing and school engagement; and 3) estimating the relative genetic and environmental influence on academic achievement beyond compulsory education and into university.

Personality and academic outcomes

Academic achievement is most parsimoniously predicted by intelligence, also referred to as general cognitive ability. General cognitive ability captures what is in common in performance on a diverse set of cognitive abilities, such as memory, reasoning, spatial ability and processing speed (Spearman, 1904). It has been shown to explain substantial variance in achievement throughout development, including childhood (Bartels et al., 2002; Spinath, Spinath, Harlaar, & Plomin, 2006), secondary school (Deary, Strand, Smith, & Fernandes, 2007; Laidra, Pullmann, & Allik, 2007) and university (Farsides & Woodfield, 2003; Ridgell & Lounsbury, 2004). However, general cognitive ability does not tell the whole story when it comes to achievement. Achievement is also influenced by a multitude of psychological factors, including personality (Conard, 2006; Farsides & Woodfield, 2003; Poropat, 2009) and motivation (Spinath et al., 2006).

Links between these psychological factors and achievement have traditionally been explored in terms of environmental associations. For example, the positive correlation between conscientiousness and achievement can be partly explained by time spent on homework (Trautwein & Lüdtke, 2007). However, there is also evidence for the influence of *genetics* on psychological factors and their associations with academic achievement. Take the previous example of conscientiousness and achievement – another explanation is that the same genes involved in differences in conscientiousness are also contributing to differences in achievement. Indeed, behavioural genetic studies have shown that inherited DNA differences account for over 50% of the covariance between measures of conscientiousness and achievement (Luciano, Wainwright, Wright, & Martin, 2006; Rimfeld, Kovas, Dale, & Plomin, 2016). Genetics also explains substantial covariance between academic achievement and other psychological traits, such as self-efficacy (Caprara, Alessandri, Di Giunta, Panerai, & Eisenberg, 2010), motivation (Gottschling, Spengler, Spinath, & Spinath, 2012) openness and neuroticism (Rimfeld et al., 2016).

The classical twin study is the most popular design in behavioural genetics for deriving estimates of heritability. This method compares identical and non-identical twins to decompose trait variance, or between-trait covariance, into additive genetic, shared environmental and non-shared environmental portions (see Methods for more information). Although twin studies have contributed immensely to our understanding of causes of individual differences, a new technique – genome-wide polygenic scores (GPS) – allows us to use DNA alone to quantify genetic effects related to personality traits and their association with academic achievement.

GPS are individual-specific scores that are created by aggregating the effects of many thousands of DNA variants across the genome, which have previously been identified through large genome-wide association (GWA) studies (see Methods for more information). One such GPS, which is based on a large GWA study of 300,000 individuals, is for educational attainment (Okbay, Beauchamp, et al., 2016). This GPS, henceforth referred to as *EduYears*, has been shown to explain variance in independent samples for a range of cognitive-based traits, such as general cognitive ability (3%) (Selzam, Krapohl, et al., 2017), reading ability (5.1%) (Selzam, Dale, et al., 2017), and educational achievement (9.1%) (Selzam, Krapohl, et al., 2017).

Although *EduYears* GPS appears to be explaining modest variance in cognitive abilities, there is less research investigating its association with personality traits. Why might we expect a GPS for educational attainment to predict significant variance in personality traits? Previous research has shown that staying in education is about more than just general cognitive ability and achievement, it is also influenced by motivation (Kuyper, Van der Werf, & Lubbers, 2000), self-efficacy (Zimmerman, Bandura, & Martinez-Pons, 1992), perseverance (Duckworth, Peterson, Matthews, & Kelly, 2007) and personality (Busato, Prins, Elshout, & Hamaker, 2000; Flouri, 2006; Fredricks, Blumenfeld, & Paris, 2004). With this in mind, we might therefore expect that the genetic variants related to educational attainment are also involved in differences in personality traits.

The work presented Chapter 2 extends previous research into the predictive power of *EduYears* GPS by using it to predict a range of personality and motivation traits. The predictive power of *EduYears* GPS is compared to two personality GPS – wellbeing (Okbay, Baselmans, et al., 2016) and neuroticism (Luciano et al., 2018) – which are created using information from GWA studies of similar sample size. Furthermore, because general cognitive ability is modestly correlated with many personality traits (Harris, 2004; Krapohl et al., 2014), this thesis tests whether associations remain after accounting for general cognitive ability. To investigate the shared aetiology of personality and academic achievement, this thesis estimates the extent to which *EduYears* GPS explains the covariance between five personality domains and educational achievement at the end of compulsory education at age 16.

School-wide environments and achievement: school type

Examination results at the end of compulsory education are of great importance. Achievement in these tests represent a major tipping point in life, opening up avenues for higher education, including university and beyond. Even for those intending to go straight into employment or training, good examination results are often a prerequisite. Considering that much of the knowledge needed to do well in these examinations is taught in school, the school environment is a likely candidate for influencing differences in student achievement. Previous research has identified many possible school environments which are thought to influence achievement, for example: teacher quality (Chetty, Friedman, & Rockoff, 2011), teacher qualifications (Clotfelter, Ladd, & Vigdor, 2010), class size (Ecalle, Magnan, & Gibert, 2006; Glass & Smith, 1979; Nye, Hedges, & Konstantopoulos, 2000), and even school start time (Carrell, Maghakian, & West, 2011; Edwards, 2012). One hotly debated school environment thought to be responsible for some of the variation in achievement is school type (Bifulco & Ladd, 2006; Clark, 2010; Marsh, 1989).

In the UK, there are three main secondary school types: state schools that are non-selective ('non-selective'), state schools that are selective ('grammar'), and fee-paying schools that are selective ('private'). Pupils attending these different school types come out with different examination results at the end of compulsory education. Indeed, 99% of grammar school students obtain top grades (A*–C grade) in English and mathematics, compared to 64% for all state-funded mainstream school students (Department for Education UK Government, 2016).

However, by design, selective schools choose their intake based on certain pupil characteristics, such as ability, achievement or, in the case of school fees, family socioeconomic status. These characteristics have been shown to correlate with academic achievement at the end of compulsory education (Deary et al., 2007; Laidra et al., 2007; Rimfeld et al., in press; Sirin, 2005). Take grammar schools for example, these schools select their intake based on an ability test at the end of primary school, and only the top achieving students are offered a place. Yet previous research has shown that general cognitive ability at this age explains almost 50% of the variance in examinations at the end of compulsory schooling at age 16 (Deary et al., 2007). Therefore, it is important to untangle the unique prediction of school type on examination results, accounting for student differences.

The relationship between the factors involved in school admission (ability, prior achievement and socioeconomic status) and later achievement is often framed environmentally. For example, parents with higher general cognitive ability may have more books in the home or facilitate intellectual discussions which, in turn, may lead to improved academic achievement of their child. However, a less frequently discussed and investigated pathway is genetics. Each of the factors that selective schools choose their intake on – ability, achievement and family socioeconomic status – show genetic influence (Knopik, Neiderhiser, DeFries, & Plomin, 2016; Krapohl & Plomin, 2016; Krapohl et al., 2014; Plomin & Deary, 2015; Trzaskowski et al., 2014; Wainwright, Wright, Luciano, Geffen, & Martin, 2005). What happens when selective schools select on heritable factors, and to what extent do these heritable factors explain the exam score difference between students attending different school types at the end of compulsory education?

To answer these questions, Chapter 3 investigates whether selecting on heritable factors, such as ability, achievement and socioeconomic status, leads to average genetic differences between students attending different school types. Furthermore, this thesis explores the unique prediction of school type to exam score differences, once heritable selection factors are considered.

School-wide environments and achievement: school quality

Another school-wide environment investigated in this thesis is school quality. School quality is not easily objectively defined since the judgement of quality involves personal values. Indeed, a school viewed as high quality by one parent, may be thought of as poor quality by another. Previous research has explored school quality from several perspectives, for example school quality rated by teacher, inspectors, or school children (Katz, 1993). The focus of this thesis is specifically on school quality as rated by independent school inspectors.

School quality is routinely assessed in eight regions across England through school inspections. These are conducted by the Office for Standards in Education, Children's Services and Skills (Ofsted), an independent organisation which reports directly to Parliament. School inspections typically last two days and involve gathering evidence on the quality of teaching, meetings with school leaders, observations of lessons, and discussions with pupils (Ofsted, 2018). Once a school inspection has taken place, the school is given an overall effectiveness rating that falls into one of four categories:

‘Inadequate’, ‘Requires improvement’, ‘Good’, or ‘Outstanding’. Ofsted publishes this result, along with a comprehensive report, on the following website: https://reports.beta.ofsted.gov.uk/.

To find out about the ethos of a school or the achievement of its pupils, parents often turn to Ofsted reports. In two parent surveys Ofsted reports were listed as the second and third most important sources of information to parents when choosing a school for their child (Ofsted, 2017; Wespieser, Durbin, & Sims, 2015). However, like school type, student allocation to schools is not always random; the best schools are often located in more affluent areas and may attract more academically motivated students. Therefore, it is important to control for these student characteristics when estimating the influence of school quality on academic achievement.

Chapter 4 explores the relationship between Ofsted ratings of a school and child-level academic achievement and wellbeing. Does a higher-rated school mean improved achievement and greater happiness? Despite Ofsted being one of the main factors influencing parental decision-making, there have been no prior studies looking at the influence of Ofsted ratings of school quality on student-level achievement or wellbeing.

Achievement beyond compulsory education

For many students, education does not end at school. Indeed, in the UK approximately 49% of 18-21 year-olds continue to university (Ford, 2017). Getting into university, and achievement at university, depends on a plethora of factors, including parental encouragement (Cabrera & La Nasa, 2000), personality (Komarraju, Karau, & Schmeck, 2009; Musgrave-Marquart, Bromley, & Dalley, 1997; Noftle & Robins, 2007) and achievement goals (Harackiewicz, Barron, Tauer, & Elliot, 2002). However, a less frequently investigated factor that may contribute to university success is genetics.

The heritability of educational achievement across development is substantial and stable (Rimfeld et al, in press). As mentioned previously, heritability estimates of achievement across development typically range from 40-60% (Baker et al., 1996; Bartels et al., 2002; Branigan et al., 2013; Kovas et al., 2007; Krapohl et al., 2014; Rimfeld et al., in press). However, despite a comprehensive body of literature examining the genetic and environmental influences on achievement in secondary school, little research has focused on the aetiology of university achievement. Chapter 5 examines the genetic influence on a range of university access and success variables, including: achievement on university entrance examinations, university attainment (getting into university or not), university quality (the position of the university in the league tables), and university achievement (degree final grade).

During secondary school in the UK, there is a standardised feature of the learning environment - the National Curriculum. Standardisation within the environment has been associated with greater influence of genetics in terms of academic achievement and intelligence (Heath et al., 1985; Samuelsson et al., 2005; Turkheimer, Haley, Waldron, d'Onofrio, & Gottesman, 2003). However, at university there is no standardised curriculum. Universities themselves are responsible for setting and often marking work. Therefore, genetic influence on achievement measures may be weaker at university.

Despite being less standardised, university also presents an opportunity for students to tailor their educational experience to a greater extent. Students decide whether they wish to attend university, what subject they want to study, which university they want to attend, whether they turn up for classes, whether they revise for exams and so on. In this way, students can choose environments based on their natural abilities, interests and aptitudes. This describes a concept, known as *gene-environment correlation*, whereby individuals select, modify and experience their environments in part based on their genotype. For example, someone who is naturally talented at music may decide to study music at university, take up several instruments and socialise with other musicians. This individual has sought out features of the environment that allow them to express their genetic potential. Gene-environment correlation opens up a new way of looking at environments as personal and chosen instead of random and brought about by uncontrollable events.

Gene-environment correlation has been shown for a range of traits assumed to be environmental, including life events (Bolinskey, Neale, Jacobson, Prescott, & Kendler, 2004; Saudino, Pedersen, Lichtenstein, McClearn, & Plomin, 1997), media use (Ayorech, von Stumm, Haworth, Davis, & Plomin, 2017) and occupational status (Fulker & Eysenck, 1979). For reviews see (Jaffee & Price, 2007; Plomin & Bergeman, 1991). The extent to which choosing to attend university and university quality are genetically influenced will be an indication of gene-environment correlation.

Chapter 5 explores the heritability of achievement into university, as well as other university success-related variables, including whether students decide to continue education to university level and the quality of the university attended. Furthermore, this chapter examines the extent to which correlations between these variables are explained by genetic and environmental factors.

In summary, the research presented in this thesis aims to explore why individuals differ in their educational outcomes. It does this by: 1) investigating the extent to which genetic variants associated with educational attainment explain differences in personality; 2) the extent to which these genetic variants explain the covariation between personality and educational achievement; 3) exploring the influence of two widely debated school environments – school type and school quality – and their unique prediction of educational achievement; and 4) clarifying the heritability of educational achievement into university, as well as other university success related variables.

METHOD

Sample

The research presented in this thesis uses data from the Twins Early Development Study (TEDS). TEDS is longitudinal UK population-based sample of twins born in England and Wales (Haworth, Davis, & Robert, 2013). Twin births between January 1994 and December 1996 were identified through birth records, and a total of 16,810 pairs of twins were recruited into the study. Since initial enrolment into the study, the families were invited to take part in data collection when twins were: 2, 3, 4, 7, 8, 9, 10, 11, 14, 16, 18 and 21 years of age. Although there has been some attrition over the years, more than 10,000 pairs remain actively involved in the study. Importantly, the sample is broadly representative of the UK population for several key traits such as socioeconomic status, ethnicity, and parental education (Haworth et al., 2013; Kovas et al., 2007).

In addition to the twin sample, there are also approximately 7,000 unrelated individuals who have been genotyped (one twin randomly in a pair). This genotyped subsample is representative of UK census data at first contact for gender, parental education, mother’s employment, and father’s employment (Selzam, Krapohl, et al., 2017).

Measures

Measures collected by TEDS

Over the past two decades, TEDS have collected data on a broad range of cognitive, behavioural and psychological traits. Data have been collected using a variety of different methods, including: paper questionnaires, telephone interviews, web testing, and mobile phone testing, and by different raters, including twins themselves, parents, and teachers (Haworth et al., 2013). A detailed description of each of the measures can be found in the relevant chapters. This thesis focuses on achievement, personality, and school environment measures collected when twins were aged 11, 16 and 21 years old. Figure 1 shows all the measures used in this thesis by age.

Measures collected by the National Pupil Database

The National Pupil Database (NPD) is a pupil-level database which holds a variety of information about students who attend schools in England https://www.gov.uk/government/collections/national-pupil-database). The NPD combines the examination results of pupils with information on pupil and school characteristics. Within the TEDS sample, 13,392 individuals gave consent for us to access their NPD records, of which 12,717 individuals were successfully matched (Smith-Woolley et al., 2018). Approximately 700 individuals who had given consent lived outside England (for example Wales or Scotland), and therefore individuals could not be matched. In the present thesis, we used pupil achievement data from the NPD to increase sample size. We compared individuals who had both achievement data from TEDS as well as NPD data and found them to be almost perfectly correlated (r = .99).

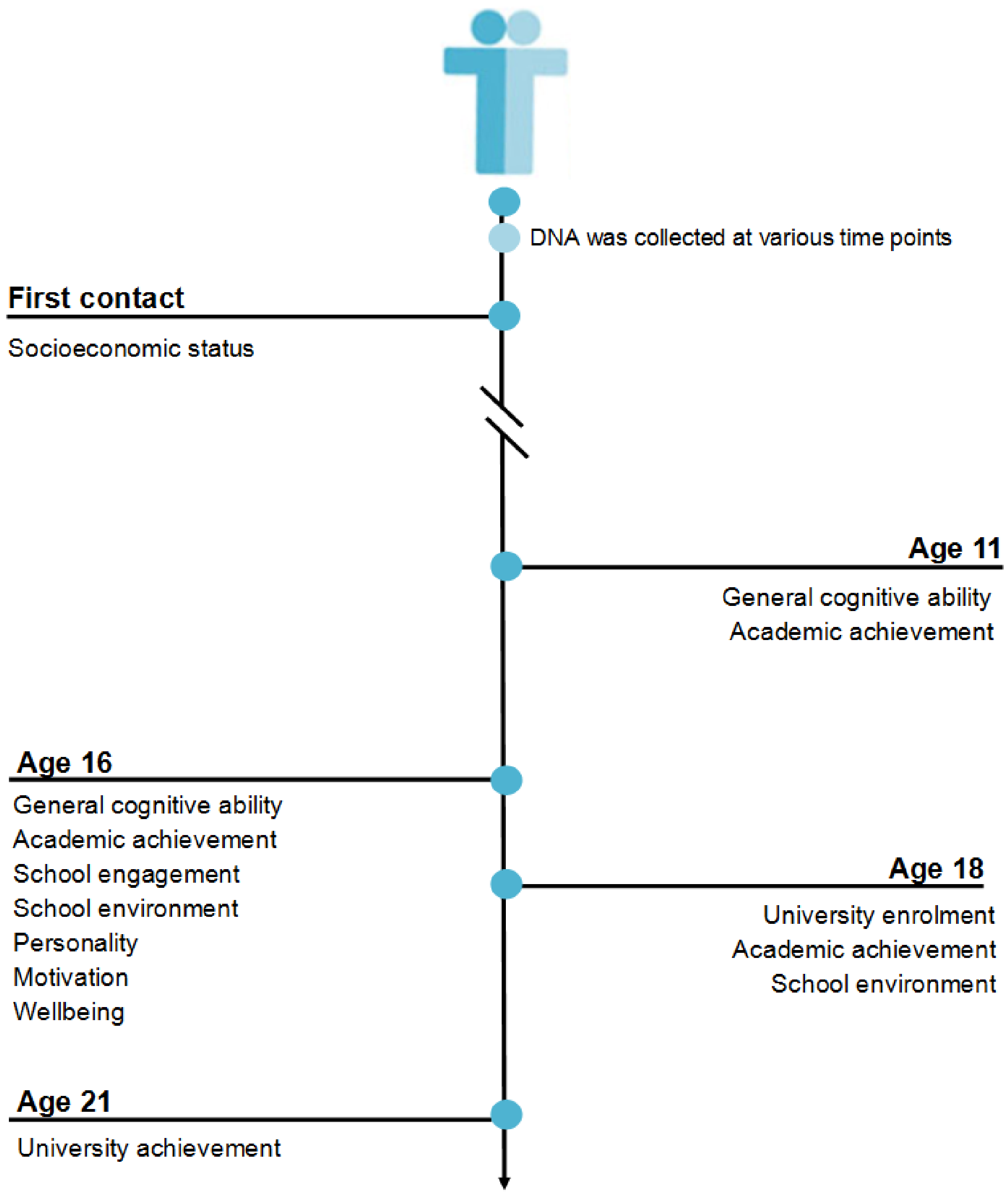


Figure 1. Measures used in this thesis by age

Twin studies

For almost a century, researchers have taken advantage of quantitative behavioural genetic designs to untangle genetic and environmental influence on traits. The most common of these designs is the classical twin study (Rijsdijk & Sham, 2002). Twin studies are made possible by the existence of two types of twin pairs: monozygotic twins (MZ) who are 100% genetically identical and dizygotic twins (DZ) who share on average 50% of their segregating genes. MZ twins are the result of a single zygote, which splits and forms two embryos, whereas DZ twins are the result of two separate zygotes – they are just like typical siblings who happen to be born at the same time. As such, while MZ twins can only be single sex, approximately half of DZ twins are same sex and half are opposite sex. Crucially, both sets of twins share their rearing environment to the same extent; both MZ and DZ twins share their uterine environment, and they grow up in the same family. Capitalising on the known genetic and known shared environmental coefficients between MZ and DZ twins, it is possible to decompose individual differences in a trait, into its variance components: additive genetics, shared environment, and non-shared environment and measurement error.

Heritability (h^2) represents the proportion of variance explained by additive genetic effects (A). These are effects from individual alleles at all loci that influence a trait and contribute to parent-offspring resemblance. The portion of variance that is not attributable to inherited DNA differences is the environment. This is defined very broadly and can include anything from the effect of socioeconomic status to parenting style. The environment can be subdivided into two further components of variance: the shared or 'common' environment (C) and the non-shared or 'unique' environment (E). The shared environment describes the variance in a trait that can be attributed to the factors making twins growing up in the same family more *similar* to one another, for example growing up in the same house or having the same childhood diet. On the other hand, non-shared environmental influences describe the portion of variance that can be attributed to environments that make twins growing up in the same family *different* from one another. Examples of non-shared environmental influences could be peer influences as a result of having different friends, or differential disease exposure. The non-shared environmental component also includes measurement error.

If MZ twins correlate higher for a trait than DZ twins, then genetic influence on that trait is inferred. On the other hand, if MZ and DZ twin correlations are similar, then the influence of the common environment is inferred.

These components can be roughly estimated using Falconer's formula:

$$h^2 = 2 * (r_{MZ} - r_{DZ})$$
$$c^2 = r_{MZ} - h^2$$
$$e^2 = 1 - h^2$$

As shown in the first equation, h^2 is estimated by doubling the difference between the DZ and MZ twin intraclass correlations. Any additional similarity between MZ twins, over and above the heritability, must be due to the shared environment (c^2), and the extent to which MZ twins do not correlate perfectly must be due to the non-shared environment or measurement error (e^2). For further information, see (Boomsma, Busjahn, & Peltonen, 2002; Knopik et al., 2016; Rijsdijk & Sham, 2002).

To accurately test the influence of genetic and environmental influences on measured traits, structural equation models are built and formally tested using maximum-likelihood estimation. This method aims to minimise the discrepancies between the fitted model and the observed data by iteratively adjusting the values of the model parameters until they best explain the observed data. Univariate twin studies have shown that virtually every complex human behaviour is at least partially heritable (Plomin, DeFries, Knopik, & Neiderhiser, 2016). A meta-analysis of the heritability of human traits based on 2,748 publications looking at 17,804 traits and with a combined sample size of 14,558,903 twin pairs estimates the average heritability of human traits to be 49% (Polderman et al., 2015).

Bivariate and multivariate models are extensions of univariate twin analysis that go beyond the estimation of genetic and environmental influences on a single trait to look at two or more traits. Specifically, bivariate twin models can address questions about the aetiology of the covariance between two traits, or they can address the genetic and environmental contributions to a single trait measured at two time points (trait stability and change). To do this, the same underlying principles are applied as with univariate designs. However, instead of comparing the MZ and DZ intraclass correlations, the cross trait cross twin correlations are compared. Cross trait cross twin correlations index the extent to which Twin 1's score on Trait 1 is predictive of Twin 2's score on Trait 2. For example, the extent to which Twin 1's height is predictive of Twin 2's weight. Similar to the univariate design, if the cross trait, cross twin correlations are higher for MZ twin pairs compared to DZ twin pairs, genetic influence on the covariance is inferred. Using structural equation modelling, these relationships are tested and A, C and E parameters are estimated.

Genetic correlation (r_G) estimates the extent to which the same genes influencing Trait 1 also influence Trait 2 and is therefore a measure of pleiotropy (one gene influencing two or more traits) or causality (one trait causing the other) (Ligthart & Boomsma, 2012). Similarly, the shared environmental correlation (r_C) measures the extent to which the same shared environments influencing Trait 1 are also operating on Trait 2. Finally, the extent to which the same non-shared environments contribute to differences in two traits is a measure of the non-shared environmental correlation (r_E). Perhaps counterintuitively, the genetic, shared environmental and non-shared environmental correlations are independent of the individual A, C and E estimates for each trait. Therefore, there may be instances in which each trait shows high heritability, however the genetic correlation between the two traits may be low and vice versa. Bivariate models can be extended to multivariate analysis in order to look at more complex data involving three or more traits, or one trait measured at three or more time points. More detailed information on the univariate and multivariate twin models can be found in Chapter 5.

Twin studies have several important assumptions to consider: 1) MZ and DZ twins share their environment to the same extent (also known as the equal environments assumption); 2) twins are representative of the population studied for the trait(s) being investigated; and 3) mating in the population is random (no assortative mating).

The equal environments assumption posits that environmentally caused similarity is roughly the same for both sets of twins. However, this assumption has been debated, with some arguing that MZ twins experience more similar environments compared to DZ twins. For example, MZ twins are more likely to play with the same childhood friends, share a bedroom, and dress more alike compared to same-sex DZ twins (Loehlin & Nichols, 2012). If this treatment leads to greater environmentally caused similarity, then MZ correlations will be increased relative to DZ correlations and therefore inflate estimates of heritability. Conversely, if MZ twins experience greater environmental differences compared to DZ twins, MZ correlations may be reduced in relation to DZ correlations and inflate shared environmental estimates. One way to test the equal environments assumption is to study twins whose perceived zygosity is different to their tested zygosity and compare these twins to twins whose perceived zygosity is the same as their tested zygosity. In this way, perceived zygosity is used to predict twin similarity. The equal environments assumption has been tested in several ways and seems reasonable for most traits, including: psychiatric illness (Hettema, Neale, & Kendler, 1995; Kendler, Neale, Kessler, Heath, & Eaves, 1993), personality

(Plomin, Willerman, & Loehlin, 1976), parenting (Kendler, Neale, Kessler, Heath, & Eaves, 1994), intelligence (Scarr & Carter-Saltzman, 1979), and physical similarity (Klump, Holly, Iacono, McGue, & Willson, 2000).

The representativeness of twin studies also deserves consideration. The results of twin studies can only be generalisable if twins are representative of the population studied. Some have argued that brain development may differ in twins compared to singleton children (Knickmeyer et al., 2011) or that twins perform worse on tests of verbal ability and IQ (Ronalds, De Stavola, & Leon, 2005). However, research has typically found support for the representativeness of twin studies, for example twins have been found to be broadly representative of the population for health (Andrew et al., 2001), personality (Johnson, Krueger, Bouchard, & McGue, 2002), psychiatric problems (Kendler, Martin, Heath, & Eaves, 1995), emotional/behavioural problems (Moilanen et al., 1999) and, crucially for this thesis, educational performance (Deary, 2006).

The final assumption of the twin study to consider concerns random mating. Random mating would hold true if all individuals were potential partners and mating was not influenced by other traits. However, non-random, or assortative mating, has been shown for many traits, including: height, personality (Glicksohn & Golan, 2001; Mascie-Taylor & Vandenberg, 1988), intelligence (Watson et al., 2004), education (Domingue, Fletcher, Conley, & Boardman, 2014) and psychiatric problems (Nordsletten et al., 2016). Assortative mating potentially leads to underestimation of heritability because children of more similar parents are likely to be more genetically similar on average. This will not affect MZ twins who are already 100% genetically identical, however DZ twins inherit some of the same genes from both parents, making them more genetically similar and thus inflating the DZ correlation.

These assumptions and limitations of the twin method are important to bear in mind when analysing and interpreting the results from twin studies.

Genome-wide polygenic scores

Unlike twin studies which infer genetic similarity, molecular genetic designs use actual measured genetic variation to study human behaviour. In the last 50 years, there has been a rapid increase in the growth in research using human molecular genetic designs (Ayorech et al., 2016). This has led scientists to dub this period the ‘genomic era’ (Guttmacher & Collins, 2003). One method which is leading the way in terms of understanding the genetic basis of human behaviour is genome-wide association

(GWA) studies. GWA studies aim to identify genetic variants (single nucleotide polymorphisms – 'SNPs') associated with variance in human behaviour, or with increased risk of a categorically defined disorders like depression (Visscher et al., 2017). Associations between SNPs and measured traits are tested using linear regression (for continuous traits) or logistic regression (for categorical traits). Because associations are tested across a million SNPs, a stringent *p*-value threshold of 5×10^{-8} is required.

Over the past decade, GWA studies have shown that there are almost no SNPs of large effect for complex human behaviour (Chabris, Lee, Cesarini, Benjamin, & Laibson, 2015; Visscher et al., 2017). Complex traits are highly polygenic and individual SNPs have very small effect (Visscher et al., 2017). This means that GWA studies require large sample sizes to detect such effects. Luckily, due to advances in genotyping, international collaborations, and population-based biobanks (Swede, Stone, & Norwood, 2007), large-scale GWA studies are now commonplace and available for psychiatric, physiological, behavioural, and cognitive traits (GWAS catalogue: https://www.ebi.ac.uk/gwas/). One of the most commonly used applications of GWA studies is to use their results to create genome-wide polygenic scores (GPS).

GPS are individual-specific scores that index genetic predisposition for increased presentation of a trait. These scores are created by taking the summary statistics from GWA studies and applying them to an independent, genotyped sample. Specifically, GWA study identified, trait-increasing alleles are summed together and weighted by their effect size (either beta coefficients or odds ratios) for each individual in the independent sample.

As with all molecular genetic analysis, it is important to account for population stratification. Population stratification refers to differences in allele frequencies which are due to ancestry rather than due to a measured trait (Freedman et al., 2004). To make this adjustment, principal components are derived from a genomic relationship matrix of the independent sample and these principal components are regressed on the GPS. Furthermore, it is also important to correctly account for any overlap between the GWA study sample and the target sample in which GPS are created (Socrates et al., 2017). If overlap is not properly accounted for then GPS estimates will be biased upwards. Once quality control has been performed, the resulting GPS can then be used just like any other variable in a dataset. For example, it can be used to explain variance in other conceptually-related traits (Chapter 2 and 5), covariance between traits (Chapter 2) or used as an individual-specific control (Chapter 3).

Unlike twin studies, which tell us about variation in a population due to genetic effects, GPS provide individual-specific scores. A powerful way to highlight this is to look at GPS between siblings. One study, using a GPS for the trait years of education found that the sibling with the higher GPS completed more years of education compared to the sibling with a lower GPS (Domingue, Belsky, Conley, Harris, & Boardman, 2015). In this way, there is potential to use GPS to provide information beyond family risk for a range of physical and mental health problems. Another advantage of GPS is that, although GWA studies need extremely large samples to detect the small effect sizes, the target samples in which GPS are created only need to be a fraction of the size.

Finally, the difference between genetic estimates from twin studies, and those from GPS, should be noted. Twin study estimates of heritability are higher because they represent the variance in a trait due to all inherited DNA differences. GPS, on the other hand, are limited to estimating only additive genetic variance explained by common SNPs found on DNA chips in large samples. Furthermore, because GPS are created using the summary statistics from GWA studies conducted with largely European samples, results will not necessarily replicate in non-European samples. They are also limited by the target GWA study's power to detect small effects. Indeed, as GWA study sample sizes have increased, so too has GPS prediction (Plomin & von Stumm, 2018). Considering the differences between genetic estimates from twin studies and those from GPS, it is to be expected that GPS prediction will be small in comparison to the sum total of genetic effects predicted by twin studies.

OVERVIEW OF THE THESIS

This thesis presents work conducted to increase understanding of the genetic and environmental contributions to academic achievement at secondary school and beyond. The aims of this thesis are to: 1) explore the contribution of attainment-related genetic variants (*EduYears* GPS) to the prediction of five personality domains, as well as their contribution to the covariance with educational achievement; 2) investigate the influence of school environments – school-type and school quality – on educational achievement, accounting for a range of genetically influenced factors; and 3) estimate the relative genetic and environmental influences on academic achievement beyond compulsory education and into university.

Academic achievement appears highly heritable across development, including at the end of compulsory schooling when children are aged 16. This high heritability is the result of many genetically influenced traits, including personality, motivation and the school environment - not just intelligence. Chapter 2 explores the extent to which genetic variants associated with educational attainment (*EduYears* GPS) predict differences in a range of personality and motivation traits, and compares this prediction to two personality GPS: neuroticism and wellbeing. Furthermore, it investigates to what extent *EduYears* GPS explains the covariance between personality and educational achievement. This chapter concludes with a discussion about the use of GPS and implications for future research.

Considering that children spend much of their childhood in school, the school environment is a likely candidate for explaining differences in academic achievement. Chapters 3 and 4 explore two school-wide environments thought to explain variance in academic achievement. Chapter 3 investigates the effect of attending different school types (selective or non-selective schools) on academic achievement, whereas Chapter 4 explores the influence of independently rated school quality on academic achievement, student wellbeing and school engagement. Because school allocation is non-random, for example certain school types are located in more affluent areas, or better-quality schools are often over-subscribed and are harder to get into, child characteristics should be accounted for. In both Chapters 3 and 4, child characteristics like prior ability, prior achievement and family socioeconomic status, are controlled for in order to look at the unique contribution of school environments to the prediction of academic achievement, and other educationally-relative traits.

Approximately 49% of school students opt to continue studying into university. Despite much research looking at the genetic and environmental influences on academic outcomes in compulsory education, surprisingly little research has focused on the heritability of university measures, including university enrolment, university quality, and university achievement. Chapter 5 explores the aetiology of these measures using the twin design as well as using GPS prediction. Furthermore, this chapter explores the genetic and environmental links between these measures of university access and success. The results are discussed in terms of gene-environment correlation.

Chapter 6 concludes with an exploration of the limitations of this research, together with a discussion of the implications for: 1) teachers and school leaders; 2) education policy-makers; and 3) parents. Finally, this thesis outlines possible future directions for research.

References:

Allegrini, A. G., Selzam, S., Rimfeld, K., Von Stumm, S., Pingault, J.-B., & Plomin, R. (in press). Genomic prediction of cognitive traits in childhood and adolescence. *bioRxiv*.

Andrew, T., Hart, D. J., Snieder, H., de Lange, M., Spector, T. D., & MacGregor, A. J. (2001). Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Research and Human Genetics, 4*(6), 464-477.

Ayorech, Z., Selzam, S., Smith-Woolley, E., Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). Publication trends over 55 years of behavioral genetic research. *Behavior genetics, 46*(5), 603-607.

Ayorech, Z., von Stumm, S., Haworth, C. M., Davis, O. S., & Plomin, R. (2017). Personalized media: a genetically informative investigation of individual differences in online media use. *PloS one, 12*(1), e0168895.

Baker, L. A., Treloar, S. A., Reynolds, C. A., Heath, A. C., & Martin, N. G. (1996). Genetics of educational attainment in Australian twins: sex differences and secular changes. *Behavior genetics, 26*(2), 89-102.

Bartels, M., Rietveld, M. J., Van Baal, G. C. M., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research and Human Genetics, 5*(6), 544-553.

Belfield, C., Crawford, C., & Sibieta, L. (2017). *Long-run comparisons of spending per pupil across different stages of education*. London, UK: Institute for Fiscal Studies.

Bifulco, R., & Ladd, H. F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy, 1*(1), 50-90.

Bolinskey, P. K., Neale, M. C., Jacobson, K. C., Prescott, C. A., & Kendler, K. S. (2004). Sources of individual differences in stressful life event exposure in male and female twins. *Twin Research and Human Genetics, 7*(1), 33-38.

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics, 3*(11), 872-882.

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social forces, 92*(1), 109-140.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29*(6), 1057-1068.

Cabrera, A. F., & La Nasa, S. M. (2000). Understanding the college-choice process. *New directions for institutional research, 2000*(107), 5-22.

Caprara, G. V., Alessandri, G., Di Giunta, L., Panerai, L., & Eisenberg, N. (2010). The contribution of agreeableness and self-efficacy beliefs to prosociality. *European Journal of personality, 24*(1), 36-55.

Carrell, S. E., Maghakian, T., & West, J. E. (2011). A's from Zzzz's? The causal effect of school start time on the academic achievement of adolescents. *American Economic Journal: Economic Policy, 3*(3), 62-81.

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science, 24*(4), 304-312.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood*. Cambridge, Massachusetts: National Bureau of Economic Research.

Clark, D. (2010). Selective schools and academic achievement. *The BE Journal of Economic Analysis & Policy, 10*(1).

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources, 45*(3), 655-681.

Conard, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality, 40*(3), 339-346.

Deary, I. J. (2006). Educational performance in twins. *British Medical Journal, 333*, 1080-1081.

Deary, I. J., & Johnson, W. (2010). Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. *International journal of epidemiology, 39*(5), 1362-1369.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13-21.

Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M., & Boardman, J. D. (2015). Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health. *AERA open, 1*(3), 2332858415599972.

Domingue, B. W., Fletcher, J., Conley, D., & Boardman, J. D. (2014). Genetic and educational assortative mating among US adults. *Proceedings of the National Academy of Sciences*, 201321426.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology, 92*(6), 1087-1101.

Ecalle, J., Magnan, A., & Gibert, F. (2006). Class size effects on literacy skills and literacy interest in first grade: A large-scale investigation. *Journal of school psychology, 44*(3), 191-209.

Edwards, F. (2012). Early to rise? The effect of daily start times on academic performance. *Economics of Education Review, 31*(6), 970-983.

Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual Differences, 34*(7), 1225-1243.

Flouri, E. (2006). Parental interest in children's education, children's self-esteem and locus of control, and later educational attainment: Twenty-six year follow-up of the 1970 British Birth Cohort. *British Journal of Educational Psychology, 76*(1), 41-55.

Ford, D. (2017). *Participation Rates In Higher Education: Academic Years 2006/2007 – 2015/2016* (Vol. SFR47/2016). London, UK: Department for Education.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research, 74*(1), 59-109.

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., . . . Pato, C. N. (2004). Assessing the impact of population stratification on genetic association studies. *Nature genetics, 36*(4), 388-393.

Fulker, D. W., & Eysenck, H. (1979). Nature and nurture: Heredity *The structure and measurement of intelligence* (pp. 102-132): Springer.

Gathmann, C., Jürges, H., & Reinhold, S. (2015). Compulsory schooling reforms, education and mortality in twentieth century Europe. *Social Science & Medicine, 127*, 74-82.

Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2-16.

Glicksohn, J., & Golan, H. (2001). Personality, cognitive style and assortative mating. *Personality and Individual Differences, 30*(7), 1199-1209.

Gottschling, J., Spengler, M., Spinath, B., & Spinath, F. M. (2012). The prediction of school achievement from a behavior genetic perspective: Results from the German twin study on Cognitive Ability, Self-Reported Motivation, and School Achievement (CoSMoS). *Personality and Individual Differences, 53*(4), 381-386.

Guttmacher, A. E., & Collins, F. S. (2003). Welcome to the genomic era. *The New England Journal of Medicine, 349*(10), 996-998.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology, 94*(3), 562-575.

Harris, J. A. (2004). Measured intelligence, achievement, openness to experience, and creativity. *Personality and Individual Differences, 36*(4), 913-929.

Haworth, C., Davis, O., & Robert, P. (2013). Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics, 16*(1), 117-125.

Heath, A. C., Berg, K., Eaves, L. J., Solaas, M. H., Corey, L. A., Sundet, J., . . . Nance, W. E. (1985). Education policy and the heritability of educational attainment. *Nature, 314*(6013), 734-736.

Hettema, J. M., Neale, M. C., & Kendler, K. S. (1995). Physical similarity and the equal-environment assumption in twin studies of psychiatric disorders. *Behavior genetics, 25*(4), 327-335.

Jaffee, S. R., & Price, T. S. (2007). Gene–environment correlations: A review of the evidence and implications for prevention of mental illness: Nature Publishing Group.

Johnson, W., Krueger, R. F., Bouchard, T. J., & McGue, M. (2002). The personalities of twins: Just ordinary folks. *Twin Research and Human Genetics, 5*(2), 125-131.

Katz, L. G. (1993). Five Perspectives on Quality in Early Childhood Programs. Perspectives from ERIC/EECE: A Monograph Series, No. 1.

Kendler, K. S., Martin, N. G., Heath, A. C., & Eaves, L. J. (1995). Self-report psychiatric symptoms in twins and their nontwin relatives: Are twins different? *American Journal of Medical Genetics, 60*(6), 588-591.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior genetics, 23*(1), 21-27.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1994). Parental treatment and the equal environment assumption in twin studies of psychiatric illness. *Psychological Medicine, 24*(3), 579-590.

Klump, K. L., Holly, A., Iacono, W. G., McGue, M., & Willson, L. E. (2000). Physical similarity and twin resemblance for eating attitudes and behaviors: a test of the equal environments assumption. *Behavior genetics, 30*(1), 51-58.

Knickmeyer, R. C., Kang, C., Woolson, S., Smith, J. K., Hamer, R. M., Lin, W., . . . Gilmore, J. H. (2011). Twin-singleton differences in neonatal brain structure. *Twin Research and Human Genetics, 14*(3), 268-276.

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). *Behavioral genetics*. New York: Worth.

Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and individual differences, 19*(1), 47-52.

Kovas, Y., Haworth, C., Dale, P. S., & Robert, P. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for research in Child Development, 72*(3), vii, 1-144.

Krapohl, E., & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry, 21*(3), 437-443.

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., . . . Dale, P. S. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences, 111*(42), 15273-15278.

Kuyper, H., Van der Werf, M., & Lubbers, M. (2000). Motivation, meta-cognition and self-regulation as predictors of long term educational attainment. *Educational Research and Evaluation, 6*(3), 181-205.

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences, 42*(3), 441-451.

Ligthart, L., & Boomsma, D. I. (2012). Causes of comorbidity: pleiotropy or causality? Shared genetic and environmental influences on migraine and neuroticism. *Twin Research and Human Genetics, 15*(2), 158-165.

Loehlin, J. C., & Nichols, R. C. (2012). *Heredity, environment, and personality: A study of 850 sets of twins*: University of Texas Press.

Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T.-K., Shirali, M., . . . Fawns-Ritchie, C. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature genetics, 50*(1), 6-11.

Luciano, M., Wainwright, M. A., Wright, M. J., & Martin, N. G. (2006). The heritability of conscientiousness facets and their relationship to IQ and academic achievement. *Personality and Individual Differences, 40*(6), 1189-1199.

Marsh, H. W. (1989). Effects of attending single-sex and coeducational high schools on achievement, attitudes, behaviors, and sex differences. *Journal of Educational Psychology, 81*(1), 70-85.

Mascie-Taylor, C. N., & Vandenberg, S. G. (1988). Assortative mating for IQ and personality due to propinquity and personal preference. *Behavior genetics, 18*(3), 339-345.

Moilanen, I., Linna, S.-L., Ebeling, H., Kumpulainen, K., Tamminen, T., Piha, J., & Almqvist, F. (1999). Are twins' behavioural/emotional problems different from singletons'? *European child & adolescent psychiatry, 8*(4), 62-67.

Musgrave-Marquart, D., Bromley, S. P., & Dalley, M. B. (1997). Personality, academic attribution, and substance use as predictors of academic achievement in college students. *Journal of Social Behavior and Personality, 12*(2), 501-511.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: big five correlates of GPA and SAT scores. *Journal of personality and social psychology, 93*(1), 116-130.

Nordsletten, A. E., Larsson, H., Crowley, J. J., Almqvist, C., Lichtenstein, P., & Mataix-Cols, D. (2016). Patterns of nonrandom mating within and across 11 major psychiatric disorders. *JAMA Psychiatry, 73*(4), 354-361.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*(1), 123-151.

Ofsted. (2017). *Annual parents survey 2017: parents' awareness and perceptions of Ofsted.* London: Office for Standards in Education, Children's Services and Skills.

Ofsted. (2018). *School inspection handbook*. London, UK: Office for Standards in Education, Children's Services and Skills.

Okbay, A., Baselmans, B. M., De Neve, J.-E., Turley, P., Nivard, M. G., Fontana, M. A., . . . Derringer, J. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics, 48*(6), 624-633.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Meddens, S. F. W. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539-542.

Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic perspectives, 25*(1), 159-184.

Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influence on "environmental" measures. *Behavioral and Brain Sciences, 14*(3), 373-386.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular Psychiatry, 20*(1), 98-108.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science, 11*(1), 3-23.

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*(3), 148-159.

Plomin, R., Willerman, L., & Loehlin, J. C. (1976). Resemblance in appearance and the equal environments assumption in twin studies of personality traits. *Behavior genetics, 6*(1), 43-52.

Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics, 47*(7), 702-709.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322-338.

Ridgell, S. D., & Lounsbury, J. W. (2004). Predicting academic success: General intelligence, Big Five personality traits, and work drive. *College Student Journal, 38*(4), 607-619.

Rijsdijk, F. V., & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in bioinformatics, 3*(2), 119-133.

Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True grit and genetics: Predicting academic achievement from personality. *Journal of personality and social psychology, 111*(5), 780-789.

Rimfeld, K., Malanchini, M., Krapohl, E., Hannighan, L., Dale, P., & Plomin, R. (in press). The stability of educational achievement across school years is largely explained by genetic factors. *npj Science of Learning*.

Ronalds, G. A., De Stavola, B. L., & Leon, D. A. (2005). The cognitive cost of being a twin: evidence from comparisons within families in the Aberdeen children of the 1950s cohort study. *BMJ, 331*(7528), 1306-1311.

Rutter, M. (1982). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge: Harvard University Press.

Samuelsson, S., Byrne, B., Quain, P., Wadsworth, S., Corley, R., DeFries, J. C., . . . Olson, R. (2005). Environmental and genetic influences on prereading skills in Australia, Scandinavia, and the United States. *Journal of Educational Psychology, 97*(4), 705-722.

Saudino, K. J., Pedersen, N. L., Lichtenstein, P., McClearn, G. E., & Plomin, R. (1997). Can personality explain genetic influences on life events? *Journal of personality and social psychology, 72*(1), 196-206.

Scarr, S., & Carter-Saltzman, L. (1979). Twin method: Defense of a critical assumption. *Behavior genetics, 9*(6), 527-542.

Selzam, S., Dale, P. S., Wagner, R. K., DeFries, J. C., Cederlöf, M., O'Reilly, P. F., . . . Plomin, R. (2017). Genome-wide polygenic scores predict reading performance throughout the school years. *Scientific Studies of Reading, 21*(4), 334-349.

Selzam, S., Krapohl, E., von Stumm, S., O'reilly, P., Rimfeld, K., Kovas, Y., . . . Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular Psychiatry, 22*(2), 267-272.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research, 75*(3), 417-453.

Smith-Woolley, E., Pingault, J.-B., Selzam, S., Rimfeld, K., Krapohl, E., von Stumm, S., . . . Allen, R. (2018). Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them. *npj Science of Learning, 3*(1), 3.

Socrates, A., Bond, T., Karhunen, V., Auvinen, J., Rietveld, C., Veijola, J., . . . O'Reilly, P. (2017). Polygenic risk scores applied to a single cohort reveal pleiotropy among hundreds of human phenotypes. *bioRxiv*, 203257.

Spearman, C. (1904). " General Intelligence," objectively determined and measured. *The American Journal of Psychology, 15*(2), 201-292.

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*(4), 363-374.

Swede, H., Stone, C. L., & Norwood, A. R. (2007). National population-based biobanks for genetic research. *Genetics in Medicine, 9*(3), 141-149.

Trautwein, U., & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-students differences and within-student variation. *Journal of Educational Psychology, 99*(2), 432-444.

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., . . . Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence, 42*, 83-88.

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological science, 14*(6), 623-628.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics, 101*(1), 5-22.

Wainwright, M. A., Wright, M. J., Luciano, M., Geffen, G. M., & Martin, N. G. (2005). Multivariate genetic analysis of academic skills of the Queensland core skills test and IQ highlight the importance of genetic g. *Twin Research and Human Genetics, 8*(6), 602-608.

Watson, D., Klohnen, E. C., Casillas, A., Nus Simms, E., Haig, J., & Berry, D. S. (2004). Match makers and deal breakers: Analyses of assortative mating in newlywed couples. *Journal of Personality, 72*(5), 1029-1068.

Wespieser, K., Durbin, B., & Sims, D. (2015). *School Choice: The Parent View*. Slough: National Foundation for Educational Research.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*(3), 663-676.

Chapter 2 – Polygenic score for educational attainment captures DNA variants shared between personality-related traits and educational achievement

This chapter, investigating DNA prediction of personality-related traits and their genetic link with educational achievement, has been adapted from a manuscript currently under the second round of reviews with the Journal of Personality and Social Psychology:

Smith-Woolley, E*., Selzam, S*., & Plomin, R. (2018). Polygenic score for educational attainment captures DNA variants shared between personality-related traits and educational achievement. *Journal of Personality and Social Psychology*.

* Joint first authors

Supplementary materials for this chapter, as detailed in the text, are attached as Appendix 1.

Abstract

Genome-wide polygenic scores (GPS) can be used to predict individual genetic risk and resilience. For example, a GPS for years of education (*EduYears*) explains substantial variance in several cognitive traits such as general cognitive ability, reading and educational achievement. Personality-related traits are also known to contribute to individual differences in educational achievement, however the relationship between the *EduYears* GPS and personality-related traits remains largely unexplored. Here, we test the relationship between a GPS for *EduYears*, neuroticism and wellbeing and five personality and motivation domains including motivation, openness, conscientiousness, neuroticism and agreeableness in a UK-representative sample of 6,710 individuals assessed at age 16. We find that *EduYears* GPS is significantly associated with these domains, predicting between 1% and 2% of the variance, whereas the neuroticism and wellbeing GPS only explain 0.5% and 1% of the variance in neuroticism respectively. We also find that *EduYears* GPS explains between 7% and 16% of the associations between personality/motivation traits and educational achievement at the end of compulsory education. In contrast, the neuroticism and wellbeing GPS did not significantly tag any of the covariance. These results demonstrate that genetic effects of educational attainment relate to personality-related traits, highlighting the multifaceted nature of *EduYears* GPS.

Introduction

Education is one of society's most expensive and most widely used intervention programmes. Among the member countries of the Organisation for Economic Cooperation and Development (OECD), education accounts for between 6-15% of annual gross domestic product (Organisation for Economic Co-operation and Development, 2017) and the average young person in these countries will stay in education until the age of 22 (Organisation for Economic Co-operation and Development, 2007). Given its societal value, great importance is placed on succeeding in education, both in terms of educational attainment (education level) and education achievement (education grade).

For a century, psychologists have attempted to unravel the major predictors of individual differences in educational success. Early work showed that 'cognitive capacity' played a substantial role in education performance (Binet & Simon, 1916), a term that now many refer to as general cognitive ability or 'g'. However, it did not tell the whole story. Around the same time, Webb (1915) proposed that in addition to g, academic performance was also influenced by a ‘w’ or ‘will’ factor, representing drive or motivation (Webb, 1915). This led the way for ‘psychological’ explanations of educational success. Most now accept an interactive model of academic performance that comprises both what a person *can do* (general cognitive ability) and *how* a person will do it (personality, motivation and other psychosocial influences).

One important factor influencing both the *can* and the *how* is genetics. Inherited DNA differences play a substantial role in general cognitive ability and psychological factors such as personality and motivation, explaining up to 60% of the individual differences in these traits (Krapohl et al., 2014). Furthermore, genetics accounts for most of the link between these factors and academic achievement (Krapohl et al., 2014). These heritability estimates are typically derived from twin studies, which compare the relative similarities between identical (monozygotic; ‘MZ’) and fraternal (dizygotic; ‘DZ’) twins (Knopik, Neiderhiser, DeFries, & Plomin, 2017). However, in this study, we use a new method – genome-wide polygenic scoring – to predict a broad range of psychological traits such as personality, motivation and psychosocial factors directly from DNA. Furthermore, we estimate the role of DNA in the association between these traits and academic achievement.

General cognitive ability and educational performance

Educational achievement represents a cumulative process of acquiring many skills, gradually over time. It is influenced by a multitude of different factors including cognitive and psychological traits (Krapohl et al., 2014). One of the best predictors of educational achievement is general cognitive ability, which captures the communalities between a diverse set of cognitive measures, such as memory, verbal-reasoning and non-verbal reasoning (Plomin & Deary, 2015). General cognitive ability is the most powerful and parsimonious predictor of academic achievement across age; it is moderately correlated with academic achievement at age 9 (r = 0.44-0.49) (Spinath, Spinath, Harlaar, & Plomin, 2006), highly correlated with Scholastic Assessment Test (SAT) in students aged 14-21 (r = .82) (Frey & Detterman, 2004) and with school performance at the end of compulsory education at age 16 (r = 0.81) (Deary, Strand, Smith, & Fernandes, 2007). Furthermore, it is moderately correlated with university achievement (r = .48) (Frey & Detterman, 2004). Despite the strong links between general cognitive ability and educational achievement, the correlation is not at unity. Although general cognitive ability explains more than half of the variance in academic performance (Deary et al., 2007), much of the variance remains unexplained. Therefore, it is important to consider other explanatory factors influencing educational performance.

Personality and educational performance

Probably the most widely researched psychological correlates of educational performance are personality traits, namely dimensions of the Five-Factor Model (FFM) (McCrae & Costa, 1987). The FFM comprises: Conscientiousness (dependability and drive to achieve), Extraversion (sociability and activity), Openness to Experience (curiosity and broadmindedness), Agreeableness (warmth and friendliness) and Neuroticism (stress and anxiety). These broad 'super-traits' have been linked both positively (conscientiousness, openness and agreeableness) and negatively (neuroticism and extraversion) to academic performance (Busato, Prins, Elshout, & Hamaker, 2000; Chamorro-Premuzic & Furnham, 2003; Conard, 2006; O'Connor & Paunonen, 2007; Petrides, Chamorro-Premuzic, Frederickson, & Furnham, 2005; Poropat, 2009; Richardson, Abraham, & Bond, 2012). In addition, their underlying primary traits (most notably dutifulness, achievement-striving and anxiety) have also been associated with differences in academic performance (Chamorro-Premuzic & Furnham, 2003).

Many studies have explored the reasons for observed associations between FFM dimensions and academic performance – both in terms of attainment and achievement. Conscientiousness is comparable to the 'w' factor described by Webb (1915) and has been linked to academic effort (Trautwein, Lüdtke, Roberts, Schnyder, & Niggli, 2009) through time spent on homework (Trautwein & Lüdtke, 2007) and time use efficiency (Kelly & Johnson, 2005). It has been shown to predict academic performance at high-school (Heaven & Ciarrochi, 2008; Laidra, Pullmann, & Allik, 2007), undergraduate (Chamorro-Premuzic & Furnham, 2003; Conard, 2006; Wagerman & Funder, 2007) and even postgraduate level (Hirschberg & Itkin, 1978). Agreeableness and Openness have also been linked to academic performance: Agreeableness through following teacher instructions and learning style (Busato, Prins, Elshout, & Hamaker, 1998) and Openness through critical thinking (Bidjerano & Dai, 2007) and intelligence (Holland, Dollinger, Holland, & Macdonald, 1995; McCrae & Costa, 1997). Like Conscientiousness, Openness is also related to success in school and at university, showing positive correlations with undergraduate and postgraduate examination scores (Geramian, Mashayekhi, & Ninggal, 2012; Laidra et al., 2007). In contrast, Neuroticism and Extraversion have been negatively linked to academic achievement; Extraversion through distractibility, sociability and problems regulating effort devoted to academic tasks (Bidjerano & Dai, 2007) and Neuroticism through stress linked with exams and poor impulse control (Zeidner & Matthews, 2000).

Because of the intercorrelations between personality traits, general cognitive ability and academic achievement, an important question to consider is how these personality traits link to achievement over and above cognitive ability. In a study looking at how the FFM dimensions link to achievement it was found that only Openness explained additional unique variance in university achievement whilst controlling for general cognitive ability (Farsides & Woodfield, 2003). Conscientiousness has also been linked to achievement over and above general cognitive ability. For example, it was demonstrated that Conscientiousness was largely independent of intelligence and that when academic achievement at secondary school was accounted for, conscientiousness continued to predict achievement at university (Poropat, 2009). This is in line with another study, also showing that once prior achievement on SATs were accounted for, Conscientiousness incrementally predicted later achievement (Conard, 2006). However, there have been few studies looking at personality and general cognitive ability concurrently at secondary school level.

Motivation and educational performance

In addition to personality dimensions, other psychological explanations of academic performance have been put forward. In a systematic review of psychological traits, Richardson and colleagues (Richardson et al., 2012) suggest five domains influencing educational success: 1) personality traits; 2) motivational factors; 3) self-regulatory strategies; 4) student's approaches to learning; and 5) psychosocial influences. Although the authors note that these domains are 'conceptually overlapping', they argue that it is important to consider a wide variety of 'non-intellective' factors when predicting academic performance.

One of these non-intellective factors, which has consistently been linked to academic performance, is motivation. Although aspects of motivation correlate moderately with the FFM dimensions, for example extraversion (positively) and neuroticism (negatively) (Komarraju & Karau, 2005), many argue that elements of motivation, such as self-efficacy beliefs, may influence achievement over and above these dimensions (Caprara, Vecchione, Alessandri, Gerbino, & Barbaranelli, 2011).

Self-efficacy beliefs are an individual's beliefs about their capabilities to produce effects (Bandura, 1997). Self-efficacy and related traits, such as self-perceived ability, engagement and academic self-concept are important constructs that help to explain students' learning and progress (Multon, Brown, & Lent, 1991; Schunk, 1989). In one study, specifically looking at math self-efficacy and self-concept (Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014), moderate correlations with achievement in maths and science were found (*r* = .17 - .58), and math self-efficacy was also a significant predictor of university entry. Similarly to personality dimensions, self-efficacy beliefs have also been shown to predict academic achievement over and above general cognitive ability: self-perceptions of ability explained an extra 8% of the variance in math achievement and 9% in English achievement at age 9 after accounting for general cognitive ability (Spinath et al., 2006).

Heritability of psychological traits

When thinking about the causes of individual differences in psychological traits, such as personality or motivation, many start with obvious environmental candidates such as schools and classrooms. However, one often overlooked factor, which has been shown to explain substantial variance (i.e. individual differences) in personality, is genetics.

Family studies allow researchers to estimate the proportion of variance in a trait due to genetic and environmental factors by comparing individuals with varying levels of genetic relatedness. In the case of twin studies, monozygotic (MZ) twins who share 100% of their DNA are compared to dizygotic (DZ) twins who share on average 50% of the DNA that varies between people. Because both sets of twins grow up in equally similar environments (Derks, Dolan, & Boomsma, 2006; Kendler, Neale, Kessler, Heath, & Eaves, 1993), the influence of genetics and the environmental on traits can be untangled: if MZ twins correlate higher for a trait than DZ twins, then genetic influence is inferred. This genetic influence is also referred to as heritability, which describes the proportion of individual differences in a trait that is the result of inherited DNA differences between individuals.

Heritability of psychological traits varies by age and by trait. However, a recent meta-analysis of 2,748 twin studies (Polderman et al., 2015) showed that for temperament and personality traits, the average heritability was 47%. In line with this estimate, one study looking at the heritability of specific psychological domains such as personality, motivation and psychosocial factors found that at age 16, heritability ranged from 35% for wellbeing to 40% for self-efficacy and up to 46% for aspects of personality (Krapohl et al., 2014). Furthermore, in the same study, it was found that inherited DNA differences explained up to 92% of the observed correlation between psychological factors and academic achievement. This suggests that genetic factors are driving much of the association between psychological factors and academic achievement at this age.

Using DNA to predict psychological traits

In addition to family studies, such as twin designs, DNA-based methods have also shed light on genetic influence on psychological traits. Genome-wide association (GWA) studies test associations between millions of known DNA variants, called single nucleotide polymorphisms (SNPs), and phenotypic traits in large samples comprising thousands of individuals. GWA studies have shown that effect sizes between individual SNPs and complex traits are usually very small, with single SNPs generally explaining less than 0.1% of the variance each (Gratten, Wray, Keller, & Visscher, 2014). However, because these genetic effects are additive, more phenotypic variance can be explained when considering these SNPs jointly (The International Schizophrenia Consortium, 2009). By summing up the number of trait-increasing alleles, which are weighted by the GWA SNP effect sizes across thousands of SNPs, it is possible to generate a genetic score for each individual in an independent sample. These genetic scores, referred to as genome-wide polygenic scores (GPS), allow DNA-based prediction for any complex trait.

One of the largest published GWA studies for a behavioural trait is years of education (*EduYears*) (Okbay et al., 2016a; Rietveld et al., 2013b). This study, which had a sample size of ~300,000 adults, tested associations between SNPs and total years in education. Using the results from this study, indicating which SNPs are associated with years of education, as well as the effect size of each association, it is possible to create GPS in an independent, genotyped sample. Genome-wide polygenic scores for *EduYears* have been shown to explain 3% of the variance in the target trait – years of education, 3.5% of the variance in general cognitive ability, up to 5.1% in reading ability and 9.1% in educational achievement at 16 (Selzam, Dale, et al., 2017; Selzam, Krapohl, et al., 2017).

Although ‘cognitive’ GPS such as years of education and intelligence appear to be explaining variance in their target traits, and related traits such as achievement (Plomin & von Stumm, 2018), personality GPS have been less predictive. For example, a GPS for wellbeing explains 0.9% of the variance in wellbeing and 0.7% in neuroticism (Okbay et al., 2016). In the current study, we sought to investigate whether a polygenic score for years of education could predict variance in a range of personality and motivation domains, how this prediction compared to personality polygenic score prediction, and whether personality-related polygenic scores relate to educational achievement.

Why might a genome-wide polygenic score for education link to personality? Similarly to achievement, educational attainment (or the years in education), is influenced by a multitude of heritable traits including both cognitive and psychological factors (Fredricks, Blumenfeld, & Paris, 2004). So far, only one study (Mottus, Realo, Vainik, Allik, & Esko, 2017) has related *EduYears* GPS to personality traits. This study investigated the link between *EduYears* GPS and the Big Five personality traits in an Estonian sample of ~3,000 adults of a wide age range. *EduYears* GPS predicted 0.5% of the variance in Neuroticism and 1.2% in Openness to experience, suggesting that the polygenic score for educational attainment tags genetic variants that also relate to personality domains. However, so far no study has investigated links to other psychological aspects, such as the primary traits of personality, as well as motivation traits such as self-efficacy beliefs.

The present study

Given the genetic links between personality-related traits and educational achievement, the current study sought to explore these relationships further by testing the extent to which *EduYears* GPS correlated with personality and motivation domains, as well as their sub-traits. In addition, using a neuroticism GPS and wellbeing GPS, we contrasted the association between these non-cognitive GPS and educational achievement to *EduYears* GPS. We also tested whether associations remained after accounting for general cognitive ability. Finally, given previous quantitative genetics findings, we tested the extent to which the *EduYears*, neuroticism and wellbeing GPS explain the covariance between a range of personality-related traits and educational achievement at age 16.

Methods

Sample

The sampling frame for the present study was the Twins Early Development Study (TEDS) (Haworth, Davis, & Robert, 2013). TEDS includes 16,000 twin pairs born between 1994 and 1996 and followed from birth to the present day. Although there has been some attrition, approximately 10,000 twin pairs are still enrolled in the study, providing behavioral, cognitive and psychological data. The TEDS sample is representative of families with children in England and Wales and its representativeness is described in detail elsewhere (Haworth et al., 2013). The current study uses a genotyped subsample of TEDS which comprises 6,710 Caucasian individuals who are unrelated (i.e. one member of a twin pair). Written informed consent was obtained from parents before data collection.

Genotyping

Two genotyping platforms were used to genotype TEDS individuals because these genotyping efforts were separated by 5 years. AffymetrixGeneChip 6.0 SNP arrays were used to genotype 3,665 individuals at Affymetrix, Santa Clara (California, USA) based on buccal cell DNA samples. Genotypes were generated at the Wellcome Trust Sanger Institute (Hinxton, UK) as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/). Additionally, 4,649 individuals were genotyped on HumanOmniExpressExome-8v1.2 arrays at the Molecular Genetics Laboratories of the Medical Research Council Social, Genetic Developmental Psychiatry Centre, based on DNA that was extracted from saliva samples. After quality control, 525,859 SNPs remained for AffymetrixGeneChip 6.0 genotypes, and 600,034 SNPs for HumanOmniExpressExome genotypes. Imputation was performed separately on the two samples, based on the Haplotype Reference Consortium (McCarthy et al., 2016) and Minimac3 1.0.13 (Fuchsberger, Abecasis, & Hinds, 2014; Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012) before merging genotype data obtained from both platforms (for full details, see Krapohl et al., 2017). Genotypes from a total of 6,710 individuals passed quality control, including 3,093 individuals genotyped on Affymetrix and 3,617 individuals genotyped on Illumina (for more details, see Supplementary Methods S1).

Measures

GCSE. The General Certificate of Secondary Education (GCSE) is a standardised UK-based examination at the end of compulsory education at age 16. Students are required to take three core subjects: English, mathematics and science. For 4,741 genotyped individuals, these results were obtained from questionnaires sent via mail, in addition to telephone interviews with twins and their parents. We also obtained subject grades for an additional 620 genotyped participants that had missing TEDS self-reported data from the National Pupil database (NPD: https://www.gov.uk/government/collections/national-pupil-database). Written consent was given before accessing this data. The total sample included 5,361 genotyped individuals (M = 16.30 years; SD = 0.29 years). Subjects were graded from 4 (G; the minimum pass grade) to 11 (A*; the best possible grade). We used a mean of the three z-standardised compulsory subjects because other subjects are taken by only subsamples of the students. English, mathematics and science performance correlated highly with each other (r = 0.70 – 0.82). Furthermore, self-reported GCSE grades of TEDS participants show high accuracy, correlating 0.98 English and 0.99 for mathematics grades with data obtained for a subsample from the NPD.

General cognitive ability. Individuals were measured on multiple cognitive tests including verbal and non-verbal abilities at age 7 (M = 7.12, SD = 0.24, N = 3,735), 12 (M = 11.46, SD = 0.64, N = 3,492) and 16 (M = 16.47, SD = 0.28, N = 3,492). Age specific mean score composites were derived from four tests at age 7: Conceptual Grouping (McCarthy, 1972), Similarities, Vocabulary and Picture Completion (Wechsler, Golombok, & Rust, 1992); three tests at age 12: Raven's Progressive Matrices (Raven & Raven, 1998), General Knowledge (Kaplan, Fein, Kramer, Delis, & Morris, 1999) and Picture Completion (Wechsler et al., 1992) and two tests at age 16: Raven's Progressive Matrices (Raven & Raven, 1998) and Mill Hill Vocabulary test (Raven, Raven & Court, 1989). A general cognitive ability composite was created by taking the arithmetic mean of the z-standardised cognitive ability composites, requiring data to be present for at least two ages (N = 2,633).

Personality and motivation measures. We included 28 self-report measures collected at age 16 (M = 16.32 years; SD = 0.68 years) via self-reports using paper booklet (b) and web-based (w) assessment:

(w) PISA maths self-efficacy – 8 items (PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org): This scale was selected from the PISA 2000, 2003 and 2006 student questionnaires, comprising 8 items asking participants to rate how confident they feel about having to do mathematical tasks on a 4-point scale from 'Not at all confident' to 'Very confident'. For example, solving an equation like: 2(+ 3) = (x + 3)(x – 3). The total score was created by taking the mean of the 8 items, requiring at least 4 to be present. The scale has an average reliability of 0.83 across OECD countries (Ray & Margaret, 2003).

(w) PISA math interest – 3 items (PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org): This scale was selected from the PISA 2000, 2003 and 2006 student questionnaires. The scale asked participants to rate how interested they were in mathematics on a 4-point scale from 'Strongly disagree' to 'Strongly agree’. For example, rating statements such as: ‘I look forward to my mathematics lessons’. The total score was created by taking the mean of the 3 items, requiring at least 2 to be present. The mean reliability across OECD countries is 0.75 for this measure (Ray & Margaret, 2003).

(w) PISA time spent on math – 3 items (PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org): This scale was selected from the PISA 2000, 2003 and 2006 student questionnaires. The scale asked participants to rate how much time they typically spent per week studying mathematics from ‘No time’ to ‘6 hours or more’. For example, ‘Study or homework in mathematics by myself’. The total score was created by taking the mean of the 3 items, requiring at least 2 to be present. The mean reliability across OECD countries is 0.76 for this measure (Ray & Margaret, 2003).

(w) Academic self-concept – 11 items (Burden, 1998). This scale aims to assess children’s perceptions of themselves as learners and problem solvers by asking children to rate themselves on a 5 point scale from ‘Very much like me’ to ‘Not at all like me’ to statements such as ‘I know the meaning of lots of words’. The total score was created by taking the mean of the 11 items, requiring at least 5 to be present. The mean reliability across OECD countries is 0.79 for this measure (Ray & Margaret, 2003).

(w) Total attitude towards key subjects – 3 items (PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org): This scale was selected from the PISA 2000, 2003 and 2006 student questionnaires. Participants were asked to answer the question: ‘In general, how important do you think it is for you to do well in the subjects below?’ on a 4 point scale from ‘Not at all important’ to ‘Very important’ for the subjects English, mathematics and science. The total score was created by taking the mean of the 3 items, requiring at least 2 to be present. The mean reliability across OECD countries is 0.79 for this measure (Ray & Margaret, 2003).

(w) School engagement – 19 items (Appleton, Christenson, Kim, & Reschly, 2006): This scale aims to assess children’s engagement with the school environment, including teacher-student relations, control and relevance of school work, peer support and family support for learning. Participants were required to answer questions such as ‘I enjoy talking to the teachers at my school’ and ‘Students at my school respect what I have to say’ on a 4 point scale from ‘Strongly disagree’ to ‘Strongly agree’. The total score was created by taking the mean of the 19 items, requiring at least 10 to be present. The reliability of factors in this measure range from 0.76 to 0.88 (Appleton, Christenson, Kim, & Reschly, 2006).

(w) Big five personality (Extraversion, Openness, Agreeableness, Conscientiousness, Neuroticism) – 30 items (Mullins-Sweatt, Jamerson, Samuel, Olson, & Widiger, 2006): We used the subscales from this measure, tapping into Extraversion, Openness, Agreeableness, Conscientiousness and Neuroticism.

***Extraversion* – 5 items:** participants were asked to rate where they were on a scale that varied for each item. For example, for the trait 'activity' they had to rate where they were on a scale from 'vigorous, energetic, active' to 'passive, lethargic'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. Across five studies, the reliability of this dimension has been estimated to be between 0.60 - 0.76 (Mullins-Sweatt et al., 2006).

***Openness* – 5 items**: participants were asked to rate where they were on a scale that varied for each item. For example for the trait 'Fantasy' they had to rate where they were on a scale from 'dreamer, unrealistic, imaginative' to 'practical, concrete'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. Across five studies, the reliability of this dimension ranged between 0.51 - 0.69 (Mullins-Sweatt et al., 2006).

***Agreeableness* – 5 items**: For example, for the trait 'compliance' they had to rate where they were on a scale from 'docile, cooperative' to 'oppositional, combative, aggressive'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. Across five studies, the reliability of this dimension ranged between 0.56 - 0.72 (Mullins-Sweatt et al., 2006).

***Conscientiousness* – 5 items:** participants were asked to rate where they were on a scale that varied for each item. For example for the trait 'self-discipline' they had to rate where they were on a scale from 'dogged, devoted' to 'hedonistic, negligent'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. Across five studies, the reliability of this dimension ranged between 0.73 - 0.78 (Mullins-Sweatt et al., 2006).

***Neuroticism* – 5 items**: participants were asked to rate where they were on a scale that varied for each item. For example for the trait 'angry hostility' they had to rate where they were on a scale from 'angry, bitter' to 'even-tempered'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. This scale was reversed, so that higher scores meant fewer neurotic traits. Across five studies, the reliability of this dimension ranged between 0.62 - 0.69 (Mullins-Sweatt et al., 2006).

(w) Ambition – 5 items (Duckworth & Quinn, 2009): This measure required participants to rate statements such as 'I aim to be the best in the world at what I do' and 'I am ambitious' on a 5-point scale from 'Very much like me' to 'Not like me at all'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. The questionnaire from which these questions were drawn has good reliability, with reliability ranging from 0.83 - 0.84 (Duckworth & Quinn, 2009).

(w) Grit – 9 items (Duckworth & Quinn, 2009): This measure required participants to rate statements such as 'I am driven to succeed' on a 5-point scale from 'Very much like me' to 'Not like me at all'. The total score was created by taking the mean of the 9 items, requiring at least 5 to be present. The questionnaire has good reliability, with reliability ranging from 0.83 - 0.84 (Duckworth & Quinn, 2009).

(w) Curiosity - 7 items (Kashdan, Rose, & Fincham, 2004): This measure required participants to rate statements such as 'Everywhere I go, I am looking out for new things or experiences' and 'I would describe myself as someone who actively seeks as much information as I can in a new situation' on a 7-point scale from 'Strongly agree' to 'Strongly disagree'. The total score was created by taking the mean of the 7 items, requiring at least 4 to be present. Across five studies, the reliability ranged from 0.72 – 0.80 (Kashdan et al., 2004).

(w) Hopefulness – 6 items (Snyder et al., 1997): This measure required participants to rate sentences about themselves, such as: 'I think I am doing pretty well' and 'I think the things I have done in the past will help me in the future' from 'All of the time' to 'None of the time'. The total score was created by taking the mean of the 6 items, requiring at least 3 to be present. Across eight studies, reliability ranged from 0.72 to 0.86, with a median alpha of 0.77 (Snyder et al., 1997).

(b) Strengths and Difficulties Questionnaire: Behavior Problems – 20 items (Goodman, 1997)**:** A dimensional and developmental measure of child mental health for children aged 3-16 years. Children are required to answer statements on a 3 point Likert scale ('Not true'; 'Quite true'; 'Very true'). It taps into 4 domains:
Conduct problems: Derived from 5 items. Item example: 'I get very angry and often lose my temper' requiring at least half of the items to be present. This scale was reversed, so that higher scores meant fewer problems. Reliability estimates across studies range from 0.44 - 0.62 (Mieloo et al., 2012).
Hyperactivity/inattention: Derived from 5 items. Item example: 'I am easily distracted, I find it difficult to concentrate'. This subscale required at least half of the items to be

present. This scale was reversed, so that higher scores meant fewer hyperactivity/attention problems. Reliability estimates across studies range from 0.75 - 0.87 (Mieloo et al., 2012).

Peer relations: Derived from 5 items. Item example: 'I have one good friend or more'. This subscale required at least half of the items to be present. Reliability estimates across studies range from 0.40 - 0.58 (Mieloo et al., 2012).

Prosocial behaviour**:** Derived from 5 items. Item example: 'I try to be nice to other people. I care about their feelings'. This subscale required at least half of the items to be present. Reliability estimates across studies range from 0.59 - 0.82 (Mieloo et al., 2012).

(b) Strengths and Weaknesses of ADHD Symptoms and Normal Behaviour Scale – 18 items (Swanson et al., 2012): This behaviour rating scale is based on DSM-5 criteria for ADHD diagnosis measuring inattentive, hyperactive, and impulsive behaviours. Children are asked to compare themselves to other people of their age on a 7-point scale from 'Far below average' to 'Far above average':

Inattention scale: Derived from 9 items. Item example: 'I sustain attention on tasks or leisure activities'. This subscale required at least half of the items to be present. On this scale, higher scores meant better attention. The reliability for this subscale is 0.91 in one English study and 0.92 in a Spanish study, with good test re-test reliability as well (r = 0.72 and 0.49) (Lakes, Swanson, & Riggs, 2012).

Hyperactivity scale: Derived from 9 items. Item example: 'I sit still (control movement of hands/ feet)'. This subscale required at least half of the items to be present. On this scale, higher scores meant better attention. The reliability for this subscale is 0.93 in one English study and 0.95 in a Spanish study, with good test re-test reliability as well (r = 0.71 and 0.61) (Lakes et al., 2012).

(w) Gratitude - 6 items (McCullough, Emmons, & Tsang, 2002): This measure required participants to rate statements such as 'I am grateful to a wide variety of people' and 'I have so much in life to be thankful for' on a 7-point scale from 'Strongly agree' to 'Strongly disagree'. The total score was created by taking the mean of the 6 items, requiring at least 3 to be present. The reliability of this scale is 0.82 (McCullough et al., 2002).

(b) Cognitive Disorganisation for cognitive disorganization – 11 items (Mason, Linney, & Claridge, 2005): This scale, measuring poor attention and concentration requires individuals to answer 11 items by answering either 'Yes' or 'No'. For example: 'Do you frequently have difficulty in starting to do things?'; 'Do you find it difficult to keep interested in the same thing for a long time?'; 'Is it hard for you to make decisions?' A total score is derived by taking the mean of the 11 items, requiring at least 6 items to be non-missing. This scale was reversed, so that higher scores meant fewer problems. Reliability of this scale is good, with Cronbach alpha estimates of 0.77 (Mason et al., 2005).

(b) Childhood Anxiety Sensitivity Index – 18 items (Silverman, Fleisig, Rabian, & Peterson, 1991): This is a child-reported questionnaire measuring anxiety sensitivity (i.e., the belief that anxiety symptoms have negative consequences). Responses are rated on a 3-point Likert scale ('Not true' to 'Very true'). For example: 'I don't want other people to know when I feel afraid'; 'I get scared when I feel nervous'. A total score is derived by taking the mean of the 18 items, requiring at least 9 items to be non-missing. This scale was reversed so that higher scores meant participants were less anxious. Reliability of this scale has been tested in clinical and non-clinical samples, both showing good Cronbach alpha's of 0.87 (Silverman et al., 1991).

(b) Moods and Feelings Questionnaire (MFQ) Short version – 11 items (Angold, Costello, Messer, & Pickles, 1995): A brief questionnaire based on DSM-III-R criteria for depression. It is measured on a 3-point Likert scale ('Not true'; 'Quite true'; 'Very true') and includes a series of descriptive phrases regarding how the participant has been feeling or acting recently. For example: 'I felt I was no good anymore'; 'I felt lonely'; 'I hated myself'. A total score is derived by taking the mean of the 11 items, requiring at least 6 items to be non-missing. This scale was reversed so that higher scores meant participants felt fewer depressive traits. The reliability of this scale is good, for both the child version (α = 0.85) and the adult version (α = 0.87) (Angold et al., 1995).

(w) Life satisfaction – 21 items (Huebner, 1994): This measure taps into different elements of life satisfaction, such as family, school, environment and life satisfaction from friends. It is measured on a 6 point scale from 'Strongly agree' to 'Strongly disagree' and asks participants to rate statements such as: 'I enjoy being at home with my family' and 'I like where I live'. A total score is derived by taking the mean of the 21 items, requiring at least 11 items to be non-missing. The reliability of this measure is good, estimated at α = 0.92 (Huebner, 1994).

(w) Subjective happiness – 4 items (Lyubomirsky & Lepper, 1999): These questions tap into perceived happiness, asking participants to complete the sentence. For example, participants are asked to complete the sentence: 'In general, I consider myself…' and are given a 7 point scale from '…Not a very happy person' to '…A very happy person'. A total score is derived by taking the mean of the 4 items, requiring at least 2 items to be non-missing. Reliability estimates from 14 samples ranged from 0.79 – 0.94 (Lyubomirsky & Lepper, 1999).

(w) Optimism – 6 items (Scheier, Carver, & Bridges, 1994): This measure required participants to rate statements such as 'In uncertain times, I usually expect the best' and 'I'm always optimistic about my future' on a 5-point scale from 'very much like me' to 'Not like me at all'. The total score was created by taking the mean of the 6 items, requiring at least 3 to be present. The reliability of this measure is good, estimated at α = 0.82 (Scheier et al., 1994).

Supplementary Table S1 shows that for most measures, there were small but significant gender differences, and that for some measures there were small effects of age. Prior to any further analyses, all variables were corrected for the effects of gender and age using the regression method to obtain z-standardised residuals.

Due to the large number of measures and the substantial overlap across these specific measures (Supplementary Figure S1), we performed principal component analysis (PCA) as a data reduction approach (Jolliffe, 1986). This method identifies underlying dimensions (principal components) that capture more variance than a single variable itself, as indicated by an eigenvalue greater than one. Because chance covariation in the data can produce eigenvalues greater than one (Jackson, 1993), we used parallel analysis (Horn, 1965) to empirically inform the eigenvalue threshold for component retention. In parallel analysis, PCA is repeatedly applied to sets of randomly generated, uncorrelated data. These data contain the same sample parameters as the study sample, and by simulating numerous PCAs, produces a distribution of eigenvalues. If the component eigenvalue in the study sample is greater than the 95th percentile of the simulated eigenvalues, the retention of this component is justified (Oconnor, 2000). Results from parallel analysis based on our sample parameters (N = 590, based on the total number of individuals with no missing data; number of variables = 28; number of iterations = 1000) indicated the retention of five components (see Table 1 for the 95th percentile of parallel analysis eigenvalues). To guide our decision-making in creating personality domains, we performed orthogonal rotation (varimax) to obtain uncorrelated factors (Kaiser, 1958).

Using PCA, five factors emerged which accounted for 53% of the total variance (Table 1). After rotation, these five factors included 'Motivation' (e.g., academic self-concept and school engagement), ‘Openness’ (e.g., ambition and curiosity), ‘Conscientiousness’ (e.g., ’grit’ and attention), ‘Agreeableness’ (e.g., gratitude and prosocial behavior) and ‘Neuroticism’ (e.g., subjective happiness and optimism). Factor loadings are shown in Table 2.

Rather than extracting factor loadings to create personality domains for subsequent analysis, which would lead to a substantial loss of data due to listwise deletion, we created variables by taking the arithmetic mean of the standardised subscales, requiring at least half to be present. Composites based on factor loading extraction and mean composite calculation correlated highly (average $r = 0.91$). Descriptive statistics of the five personality and motivation domains and the 28 subscales are shown in Supplementary Table S1. Correlations across the 28 personality and motivation measures and polygenic scores are shown in Supplementary Figure S1.

To test whether there were any meaningful differences between those with missing and non-missing personality and motivation composites, we conducted sensitivity analysis. We assessed mean differences in socio-economic status assessed at first contact (mean composite of parental education, occupation, and maternal age at the birth of the first child), general cognitive ability and GCSE results between missing and non-missing personality and motivation composites scores. We found only small differences between those with missing and non-missing data, generally accounting for less than 1% of the phenotypic variance (see Supplementary Table S2).

Table 1. Principal Component Analysis

Component	Parallel analysis Eigenvalues (95th percentile)	Initial Eigenvalues			Rotation Sums of Squared Loadings		
		Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.480	7.192	25.684	25.684	3.902	13.936	13.936
2	1.407	2.629	9.391	35.075	3.588	12.813	26.748
3	1.353	2.128	7.600	42.675	2.834	10.120	36.868
4	1.311	1.675	5.983	48.658	2.472	8.827	45.695
5	1.274	1.316	4.699	53.356	2.145	7.661	53.356
6	1.240	1.081	3.859	57.216			
7	1.208	0.934	3.337	60.552			
8	1.176	0.902	3.222	63.774			
9	1.149	0.845	3.019	66.794			
10	1.120	0.810	2.894	69.688			
11	1.093	0.791	2.826	72.513			
12	1.072	0.730	2.609	75.122			
13	1.044	0.689	2.462	77.584			
14	1.020	0.630	2.251	79.835			
15	0.995	0.580	2.070	81.905			
16	0.974	0.539	1.924	83.829			
17	0.949	0.520	1.857	85.686			
18	0.927	0.479	1.711	87.397			
19	0.903	0.461	1.646	89.043			
20	0.879	0.430	1.536	90.578			
21	0.856	0.420	1.501	92.079			
22	0.833	0.391	1.397	93.476			
23	0.812	0.366	1.307	94.783			
24	0.786	0.355	1.270	96.053			
25	0.763	0.324	1.159	97.212			
26	0.736	0.292	1.043	98.254			
27	0.709	0.254	0.908	99.162			
28	0.677	0.235	0.838	100.000			

Table 2. Rotated item factor loadings

	Neuroticism	Conscientiousness	Openness	Motivation	Agreeableness
PISA math self-efficacy				.713	
PISA math interest				.768	
PISA Time Spent on math				.502	
Academic self-concept		.408	.464	.449	
Attitude towards key subjects				.504	
School engagement				.411	.377
Extraversion	.462		.570		
Openness			.551		
Ambition		.308	.618	.312	
Curiosity			.726		
Hopefulness	.389	.306	.522		
Conscientiousness		.605	.344		
SDQ Conduct scale (r.)		.423			.382
SDQ hyperactivity (r.)		.783			
GRIT		.576			
SWAN Hyperactivity		.704			
SWAN Inattention		.774			
Agreeableness					.689
Prosocial behaviour					.618
Gratitude	.362				.550
Neuroticism (r.)	.588		.315		
Cognitive Disorganisation	.633	.502			
CASI anxiety (r.)	.687				
MFQ (r.)	.800				
Life satisfaction	.530				.402
Subjective happiness	.704				.323
Optimism	.397		.369		
Peer problems (r.)	.631				

Note: Orthogonal (varimax) rotation was applied. (r.) = recoded so that higher scores were positive, i.e. fewer conduct problem. Only variables with factor loadings greater than 0.30 are shown

Statistical Analyses

Genome-wide polygenic scores

For the 6,710 genotyped individuals in our sample, we calculated three polygenic scores. The first was based on the summary statistics for a GWA meta-analysis for years of education on 328,918 individuals (Okbay et al., 2016a). The second and third were based on the two largest GWA meta-analyses for personality-related traits to date, neuroticism (N = 329,821) (Luciano et al., 2018) and wellbeing (N = 298,420) (Okbay et al., 2016).

The first wave of TEDS genotype samples (N = 2,148) (Trzaskowski et al., 2013) was included in the discovery sample of the wellbeing GWA meta-analysis. Therefore, we performed a statistical correction on the summary statistic effect size coefficients and *p*-values (Socrates et al., 2017) to account for the overlap between the discovery and target sample. We first replicated the genome-wide association study on wellbeing using genotypes from the 2,148 TEDS individuals that were included in the meta-analysis, following the GWA protocol applied in the discovery analysis (Okbay et al., 2016). Secondly, the obtained beta coefficients and standard errors for each SNP were then used to adjust the meta-analyses beta coefficients and standard errors. These adjusted values are analogous to the effects for each SNP if the TEDS sample would have been removed in the discovery meta-analysis (Socrates et al., 2017). Third, we calculated new *p*-values based on the adjusted beta coefficients and standard errors. The adjusted summary statistics for wellbeing were used for polygenic score calculation in the full TEDS sample.

A GPS is calculated by using information from GWA study summary statistics about the strength of association between a genetic variant and a trait, to score individuals' genotypes in independent samples such as TEDS. For each individual in TEDS, all trait-associated alleles (0,1, or 2) are counted and multiplied by their effect size (i.e. their strength of association with a trait as reported in GWA summary statistics) for each SNP. The sum of these weighted and counted alleles forms a personal genomics score for each individual. We used the software PRSice (Euesden, Lewis, & O'Reilly, 2014) to create GPS. Those SNPs that passed quality control were clumped for linkage disequilibrium by applying an r^2=0.1 cut-off within a 250-kb window. Based on the *EduYears*, neuroticism and the adjusted wellbeing summary statistics, clumping reduced the number of SNPs included for further analysis to 354,866 SNPs, 364,174

SNPs and 101,553, respectively. It is possible to calculate various GPS based on different GWA study significance thresholds for genetic variants, with a lower *p*-value threshold resulting in GPS that include a higher number of SNPs. Here, we calculated GPS for seven significance thresholds (0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5) (see Supplementary Table S3 for number of SNPs included). For our analyses, we used GPS based on a *p*-value threshold of 0.1, which includes 30,086 SNPs in *EduYears* GPS, 93,493 SNPs for the neuroticism GPS, and 30,636 SNPs for the wellbeing GPS. To control for platform effects (Affymetrix vs Illumina) and plate effects, as well as effects of population stratification, we regressed all GPS used in this study on platform and plate data, and the first ten principal components. For all subsequent analyses, we used *z*-standardized residuals.

Multiple testing was accounted for by adjusting the significance threshold by the number of comparisons performed in accordance with the Šidák correction (Sidak, 1971). A total of 54 comparisons were conducted as part of our main analyses, resulting in a corrected *p*-value threshold of $1 – 0.95^{1/54} = 9.49 \times 10^{-4}$.

To test the extent to which *EduYears* GPS can predict non-cognitive traits that are related to GCSE, we used regression analysis. In addition, we used GPSs for personality related traits (neuroticism and wellbeing) to predict GCSE results and measured personality and motivation domain. Because these traits are associated with general cognitive ability, we repeated these analyses using the residuals obtained from regressing our personality and motivation traits on general cognitive ability. Additionally, we performed multiple regression analysis to assess the relative contributions of general cognitive ability and the personality and motivation phenotypes to polygenic score variation. Finally, we calculated the extent to which each GPS accounts for the relationship between personality and motivation domains and GCSE grades using structural equation modelling. By estimating the (i) GPS effect on the personality/motivation traits and GCSE grades, (ii) the residual correlation between personality/motivation traits and GCSE results after accounting for the mutual effect of the GPS on both traits, and (iii) the total effect of the model, it is possible to calculate the extent to which a GPS explains the relationship between personality/motivation domains and GCSE results (see Supplementary Methods S1).
Structural equation modelling analyses were performed in R, as implemented in the package lavaan (Rosseel et al., 2011), which is based on maximum likelihood.

Results

Correlations between personality-related domains and academic achievement

Phenotypic correlations between academic achievement (GCSE results) and the five personality and motivation domains were examined to evaluate the strength of relationships between these measures. Pearson's correlation coefficients were statistically significant and ranged from 0.13 to 0.51 (see Supplementary Figure S1). For correlations between all primary personality and motivation traits and GCSE results, see Supplementary Figure S1.

Polygenic score prediction of personality and motivation

To test the predictive validity of the polygenic score for years of education (*EduYears* GPS) and the five personality and motivation domains that contribute to educational success, we performed association analyses. Figure 1A shows that *EduYears* GPS was a significant predictor of all personality/motivation domains but Neuroticism, which did not withstand correction for multiple testing. EduYears GPS significantly explained between 0.9% - 2% of the variance in these domains. The direction of associations indicated that higher *EduYears* GPS scores related to higher Motivation, Openness, Conscientiousness and Agreeableness. We also tested the association with GCSE grades, finding EduYears GPS predicted 7% of the variance in GCSE results. Differences in the prediction of GCSE grades by *EduYears* GPS as compared to previous reports (Selzam, Krapohl, et al., 2017) are due to differences in the construction of the GCSE grade and GPS variables and an increased target sample size.

The GPS for neuroticism significantly predicted variance in GCSE results (R^2 = 0.5%), as well as the Neuroticism composite (R^2 = 0.6%) (Figure 1A). Associations with the Conscientiousness and Motivation composite did not survive multiple testing corrections, but showed suggestive significance ($p < 0.05$). Overall, the direction of effects indicated that individuals with a higher neuroticism GPS scored lower in their GCSE results, higher in the Neuroticism composite (note the reversion of this scale), and lower in Conscientiousness and Motivation.

The wellbeing GPS predicted significant variance in the Neuroticism composite (R^2 = 1%), such that a higher wellbeing GPS related to lower Neuroticism scores. No correlation was found with GCSE results (Figure 1A), and there was only a suggestive positive association with Openness. With the exception of the Neuroticism composite, the magnitudes of the correlation coefficients between *EduYears* GPS and the phenotypic measures were at least twice as high as of those relating to the neuroticism and wellbeing GPS. Non-overlapping confidence intervals of correlation coefficients for Motivation, Agreeableness, and Openness suggests a significant difference between these estimates (see Supplementary Figure S2). In contrast, non-overlapping confidence intervals of correlation coefficients suggests that the wellbeing GPS is a better predictor of neuroticism than *EduYears* GPS. Results for all other GPS thresholds are reported in Supplementary Figures S3-5.

Controlling for general cognitive ability

General cognitive ability correlated with personality and motivation primary traits and composites, as well as GCSE grades (Supplementary Figure S1). Therefore, we corrected the composites and GCSE results for variance explained by general cognitive ability and repeated the association analyses as shown in Figure 1B. We found that *EduYears* GPS was still a significant, albeit attenuated, predictor of GCSE grades, Agreeableness and Motivation. For the neuroticism GPS, previously significant correlations did not reach the multiple-testing corrected *p*-value threshold after accounting for general cognitive ability, and the strength of associations was mostly attenuated for GCSE results. The correlation between the wellbeing GPS and the Neuroticism composite remained statistically significant, with a small reduction of the correlation coefficient effect size. These results suggest that the covariance shared between the GPS and the personality and motivation domains is partly tagged by general cognitive ability, but not solely explained by it. Attenuations were more pronounced for *EduYears* GPS associations (49%) than for the neuroticism (29%) and wellbeing GPS (11%), indicating that, as expected, the *EduYears* GPS tags more variants related to general cognitive ability.

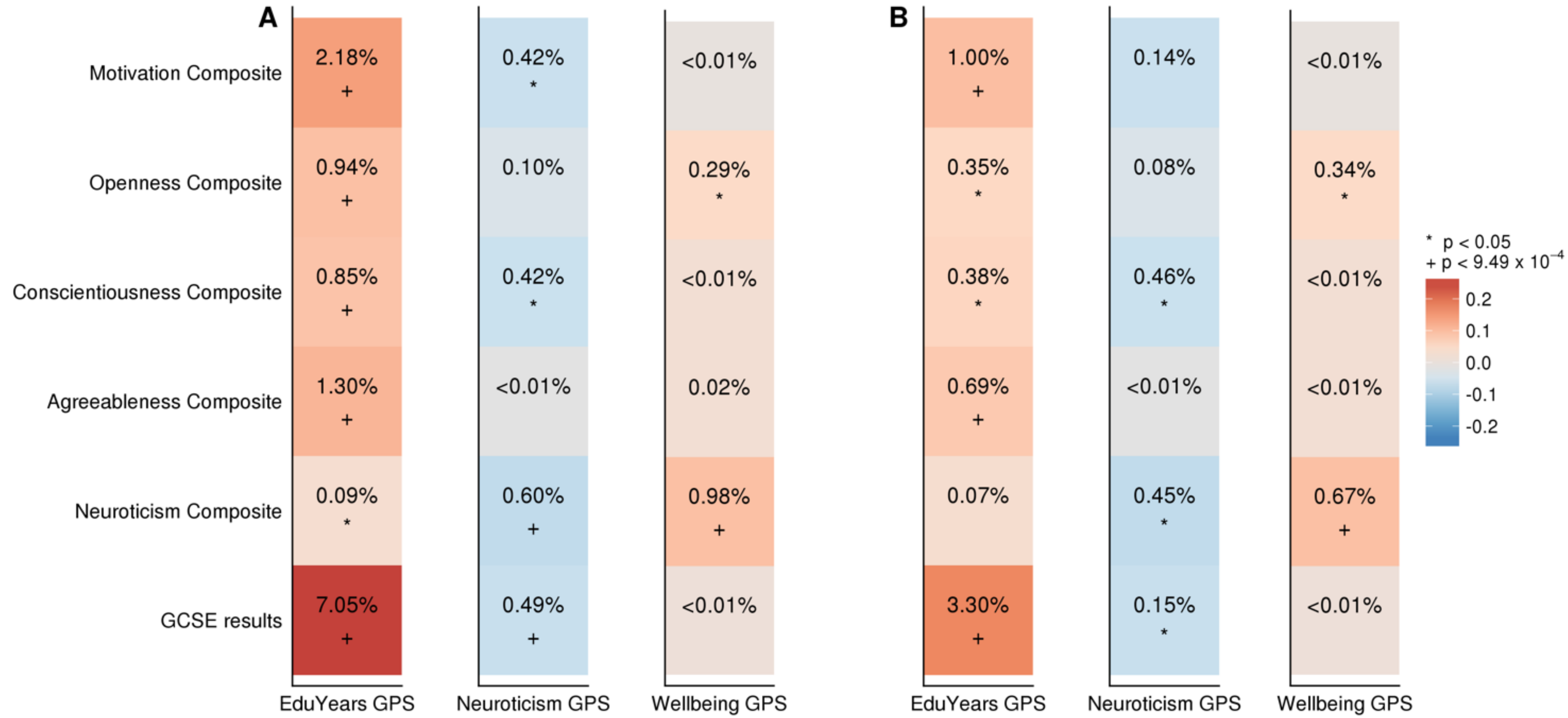


Figure 1. Genome-wide polygenic scores (GPS) as a predictor of five personality and motivation domains and GCSE (General Certificate of Secondary Education) results (**A**) before and (**B**) after accounting for general cognitive ability. The colour shading represents magnitude of Pearson Correlation coefficients and the values in each cell represent the amount of phenotypic variance explained by the polygenic scores. The scale of the Neuroticism composite was reversed, such that higher scores relate to lower neuroticism. '+' = *p*-value threshold for significance after correction for multiple testing.

Multiple regression analyses predicting polygenic scores from cognitive ability, personality and motivation

To further assess the contributions of cognitive ability and the personality/motivation domains in the polygenic score variation, we performed multiple regression analyses with the polygenic scores as dependent variables. Table 3 shows the beta coefficients for each measure in the joint prediction models. Results for Model 1 indicated that a significant proportion of variance in *EduYears* GPS was explained by the predictors (F(6,1439) = 9.96, p = 8.57 x 10^{-11}, R^2 = 0.036). Most of the effects were driven by general cognitive ability and the Motivation composite, although the Motivation composite only reached suggestive statistical significance. Both the joint prediction of the neuroticism GPS (F(6,1439) = 3.20, p = 4.03 x 10^{-3}, R^2 = 0.009), and the individual predictors, were not significant after correction for multiple testing. Similarly, the multiple regression model predicting the wellbeing GPS was not statistically significant (F(6,1439) = 3.56, p = 1.68 x 10^{-3}, R^2 = 0.011), and most of the variance was accounted for by the Neuroticism composite.

Table 3 Results from multiple regression analyses: Cognitive and personality/motivation composites predicting genome-wide polygenic scores

	Model 1: EduYears GPS			Model 2: Neuroticism GPS			Model 3: Wellbeing GPS		
Predictors	***β***	***SE***	***p***	***β***	***SE***	***p***	***β***	***SE***	***p***
General cognitive ability	0.127	0.028	6.76×10^{-6}+	-0.059	0.029	0.042*	-0.016	0.029	0.577
Motivation Composite	0.094	0.033	0.004*	-0.001	0.034	0.979	-0.029	0.034	0.389
Openness Composite	-0.022	0.032	0.487	-0.003	0.033	0.938	0.046	0.033	0.162
Conscientiousness Composite	0.002	0.034	0.962	-0.038	0.035	0.286	-0.057	0.035	0.103
Agreeableness Composite	0.067	0.032	0.036	0.028	0.033	0.400	-0.042	0.033	0.197
Neuroticism Composite	0.012	0.033	0.722	-0.083	0.035	0.016*	0.134	0.034	1.07×10^{-4}+

Note. Beta coefficients, standard errors and p-values are presented for each of the predictors in the regression models.
*$p < 0.05$, '+' = 9.49×10^{-4} (p-value threshold for significance after correction for multiple testing).

Polygenic score prediction of covariation

Because GCSE grades, *EduYears* GPS and the personality and motivation domains are all intercorrelated (Supplementary Figure S1), we tested the extent to which *EduYears* GPS accounted for the association between GCSE grades and the personality and motivation domains. Figure 2 and Table 4 show that *EduYears* GPS significantly accounted for 7% - 16% of the covariances. For comparison, we performed the same analyses using the neuroticism and wellbeing GPS. The neuroticism GPS explained 2.5% of the covariance between Neuroticism and GCSE, although the significance threshold did not survive correction for multiple testing (Figure 2; Table 3). No significant proportion of covariance was explained by the wellbeing GPS.

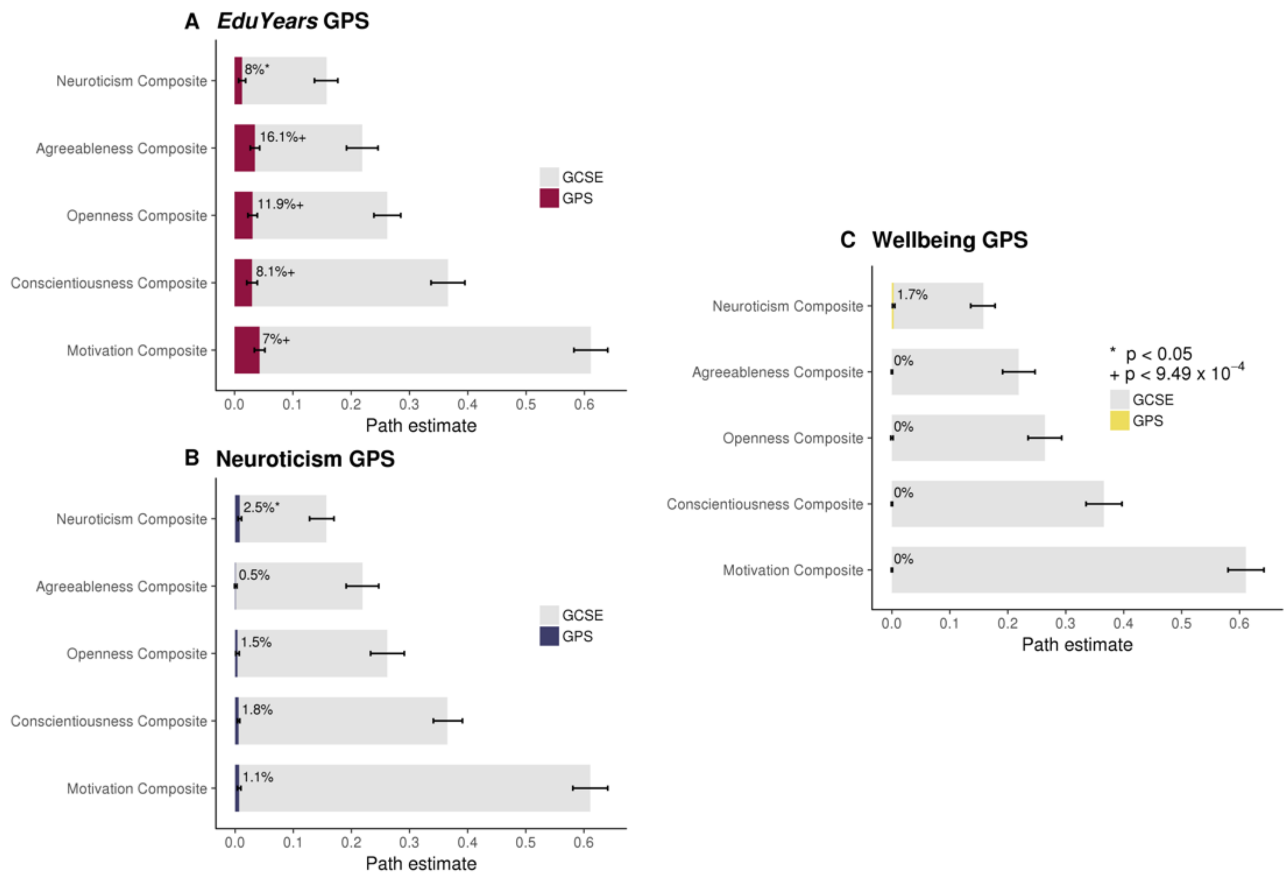


Figure 2. Standardized path estimates and standard errors for the association between GCSE (General Certificate of Secondary Education) grades and personality/motivation domains, and the proportion of these relationships accounted for by (**A**) *EduYears* GPS (genome-wide polygenic scores) (**B**) neuroticism GPS and (**C**) wellbeing GPS. Path estimates presented are estimated based on maximum likelihood (see Table 4 for all path estimates). '+' = *p*-value threshold for significance after correction for multiple testing.

Table 4 – Path estimates and standard errors

		EduYears GPS		Neuroticism GPS		Wellbeing GPS	
	parameter	***β***	***SE***	***β***	***SE***	***β***	***SE***
Motivation Composite	GPS effect	0.043^{+}	0.009	0.007	0.003	0.000	0.001
	resid cor	0.568^{+}	0.029	0.604^{+}	0.030	0.611^{+}	0.031
	total effect	0.611^{+}	0.031	0.611^{+}	0.031	0.611^{+}	0.031
	proportion	0.070^{+}	0.014	0.011	0.006	0.000	0.001
Openness Composite	GPS effect	0.031^{+}	0.008	0.004	0.003	-0.002	0.002
	resid cor	0.231^{+}	0.023	0.258^{+}	0.029	0.264^{+}	0.029
	total effect	0.262^{+}	0.029	0.262^{+}	0.029	0.262^{+}	0.029
	proportion	0.119^{+}	0.030	0.015	0.011	-0.006	0.006
Neuroticism Composite	GPS effect	0.013*	0.006	0.008*	0.003	0.003	0.002
	resid cor	0.145^{+}	0.020	0.149^{+}	0.021	0.155^{+}	0.021
	total effect	0.157^{+}	0.021	0.157^{+}	0.021	0.157^{+}	0.021
	proportion	0.080*	0.034	0.025*	0.017	0.017	0.014
Conscientiousness Composite	GPS effect	0.030^{+}	0.009	0.006	0.002	0.000	0.001
	resid cor	0.336^{+}	0.029	0.359^{+}	0.025	0.366^{+}	0.031
	total effect	0.366^{+}	0.031	0.366^{+}	0.025	0.366^{+}	0.031
	proportion	0.081^{+}	0.023	0.018	0.006	-0.001	0.002
Agreeableness Composite	GPS effect	0.035^{+}	0.008	0.001	0.002	0.000	0.001
	resid cor	0.184^{+}	0.027	0.218^{+}	0.028	0.219^{+}	0.028
	total effect	0.219^{+}	0.028	0.219^{+}	0.028	0.219^{+}	0.028
	proportion	0.161^{+}	0.036	0.005	0.011	-0.001	0.003

Note. GPS effect = effect of the genome-wide polygenic score (GPS) on both traits; resid cor = residual correlation between phenotypes after mutually adjusting for the effects of the GPS, total effect = effect accounted for by the model (resid cor + GPS effect); proportion = the proportion of the total effect that is accounted for by the GPS effect (GPS effect / total effect). *$p < 0.05$, '+' = 9.49×10^{-4} (p-value threshold for significance after correction for multiple testing).

Discussion

Our results show that a genome-wide polygenic score (GPS) for years of education (*EduYears*) predicts a number of personality and motivation domains, including Agreeableness, Openness, Conscientiousness and Motivation. We find that the educational attainment GPS is more predictive of Motivation, Openness and Agreeableness than personality-related GPS themselves, and that *EduYears* GPS explains between 7-16% of the covariance between personality and motivation domains and educational achievement at age 16. These findings suggest that DNA variants contributing to educational attainment are also important predictors of personality and motivation.

Much of the previous research using *EduYears* GPS has focused on its relationship with traditional 'cognitive' traits, such as general cognitive ability (Selzam, Krapohl, et al., 2017), reading (Selzam, Krapohl, et al., 2017) and educational outcomes (Ayorech, Krapohl, Plomin, & von Stumm, 2017). In contrast, our findings demonstrate the broad, multifaceted nature of *EduYears* GPS, which is also associated with a variety of personality and motivation traits. Indeed, we show that *EduYears* GPS significantly predicts four out of five personality and motivation domains: Motivation, Openness, Conscientiousness, and Agreeableness, explaining between 0.9% and 2.2% of the variance. However, we found that the wellbeing GPS was more predictive than *EduYears* GPS for explaining variance in Neuroticism (1% vs. 0.1%).

We find that even once we accounted for general cognitive ability, *EduYears* GPS still predicted significant variance in Agreeableness (0.7%) and Motivation (1.0%), and GCSE results as reported previously (Selzam et al., 2017a). Although correcting for general cognitive ability attenuated associations between the neuroticism and wellbeing GPS and related personality traits, the degree of attenuation was considerably smaller than for *EduYears* GPS. One possible explanation for this finding is that the *EduYears* GWA study tags more general cognitive ability related variants than the neuroticism and wellbeing GWA study, because the target trait is inherently related to individuals' general cognitive ability. Therefore, variants identified in the neuroticism and wellbeing GWA study are arguably more specific to their target trait. The findings that *EduYears* GPS is correlated with personality and motivation traits, even after accounting for general cognitive ability, are particularly interesting for two reasons. Firstly, they show that a polygenic score for years of education not only tags genetic variances associated with its target trait, but also many other traits that contribute to how long a person stays in education. And secondly, our findings illustrate that staying in education depends on more than just intelligence; many cognitive and non-cognitive genetically-influenced traits contribute to educational attainment.

In addition to showing that *EduYears* GPS explains significant variance in personality and motivation domains, we also show that it explains between 7 – 16% of the association between personality and motivation domains and educational achievement at age 16. In contrast, the neuroticism and wellbeing GPS did not significantly account for any covariance between these traits and GCSE results. As previously mentioned, a possible explanation for this finding is that GWA studies performed on personality traits may tag variants specific to the target trait, rather than capturing trait-related variants that also contribute to the development of skills important for educational achievement. In contrast, a GWA study performed on educational attainment is likely to capture genetic variants that are important contributors to many down-stream educationally relevant traits. For example, if motivation is a genetically influenced trait and an important factor for higher educational attainment, a GWA study on years of education will indirectly capture some of the genetic effects relating to motivation.

These results demonstrate the substantial genetic pleiotropy (i.e. one DNA marker affects several traits) across educational achievement and educationally relevant traits. However, it is not possible to distinguish between biological pleiotropy (i.e. one DNA marker directly affects several traits) and mediated pleiotropy (i.e. one DNA marker directly affects one trait, which then in turn affects another trait (Solovieff, Cotsapas, Lee, Purcell, & Smoller, 2013)). The findings of this study support previous twin research, showing that between 8 – 37% of the covariance between non-cognitive traits and GCSE is explained by shared genetic factors (Krapohl et al., 2014). Although the difference between the magnitudes of effect sizes from GPS and twin method results seem large, the GPS effect sizes are substantial given the limitations of the polygenic score method. In contrast to the twin method, which captures all types of genetic variation, GPS results are based on common DNA markers only. Furthermore, the predictive power of polygenic scores is directly related to the power of GWA studies to detect the small SNP effect sizes to begin with, which is one of the main difficulties faced in genetic research (Cesarini & Visscher, 2017). Due to lack of statistical power attributed to sample size and other factors, such as genotyping error or measurement error of the target phenotype, effect size estimates of specific SNPs include measurement error (Dudbridge, 2003; Mark et al., 2008; Van Der Sluis, Verhage, Posthuma, & Dolan, 2010). Therefore, these estimates are not entirely representative of the “true” genetic effect, further contributing to a downward bias of the GPS prediction.

Despite the broad range of phenotypes used within the present study, there were limitations to our measures. The first limitation concerns our personality dimension reduction analysis. Although the five dimensions that emerged from this analysis were closely aligned with the literature on personality, instead of a fifth factor for Extraversion we found a factor tapping into motivation. There are two reasons for this finding. Firstly, the measures captured by the Motivation dimension are not typically included within factor analysis of personality dimensions. These measures, (e.g. academic self-concept, self-efficacy and attitudes towards subjects) correlate with the Conscientiousness dimension (*r* = 0.18 – 0.47), as would be expected given its primary traits of ‘productive’ and ‘self-discipline’, however most of the variance is left unexplained. Secondly, the primary traits of Extraversion (e.g. ‘gregarious’, ‘excitement seeking’ and ‘warmth’) are not well covered within our measures. For these reasons, it is not surprising that a separate factor of Extraversion did not emerge and instead Extraversion loaded onto Openness.

The second limitation with our measures was the missing data. Because not everyone in our study completed all of the personality and motivation measures, there were missing data for each of our broad dimensions. To make sure that this did not affect the representativeness of the sample, we compared those with missing and non-missing data on socio-economic status, general cognitive ability and achievement at age 16. We found that missingness accounted for 1% of the variance in these outcome variables, suggesting that those with missing and non-missing data were not substantially different on these important traits.

A final limitation is that our study was limited to testing the relationships between GPS and personality domains at one age point (age 16). Future research would benefit from taking a longitudinal approach to looking at the relationship between GPS and personality traits across development. It would be interesting to explore whether genetic variants associated with years of education were differentially implicated in personality traits at different ages.

Despite the limitations to this study, it is the most comprehensive study to date investigating the link between *EduYears* GPS and personality traits. Our findings indicate a significant step towards the era of genomics, where the DNA of individuals can be used for genomic prediction, which will only become more useful as GWA studies get larger. For the time being, the current study goes some way in starting to unpick the genetic architecture of educational achievement, beyond what we have learnt from twin studies. We show that educational achievement is influenced by a multitude of different genetically influenced traits, not just general cognitive ability, and that genetic variants associated with how long students stay in education explain some of the association between personality traits and achievement. We also show that *EduYears* GPS is multi-faceted, tagging variants associated with both cognitive and personality-related traits. As GPS prediction improves based on increasing GWA study sample sizes and methodological advancements, GPS will become a powerful tool both within research, and potentially in the classroom.

Acknowledgements

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938) and the European Commission (602768; 295366). RP is supported by a Medical Research Council Professorship award (G19/2) and a European Research Council Advanced Investigator award [295366]. SS is supported by the MRC/IoPPN Excellence Award and by the EU Framework Programme 7 (602768).

Author contributions:

RP directs and received funding for the Twins Early Development Study (TEDS). RP and ESW and SS conceived the present study. ESW and SS analysed and interpreted the data with advice from RP. ESW and SS wrote the manuscript with help from RP.

Competing financial interests

The authors declare no conflict of interest.

References

Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research, 4*(5), 237-249.

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of school psychology, 44*(5), 427-445.

Ayorech, Z., Krapohl, E., Plomin, R., & von Stumm, S. (2017). Genetic Influence on Intergenerational Educational Attainment. *Psychological science*, 0956797617707270.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Macmillan.

Bidjerano, T., & Dai, D. Y. (2007). The relationship between the big-five model of personality and self-regulated learning strategies. *Learning and individual differences, 17*(1), 69-81.

Binet, A., & Simon, T. (1916). *The development of intelligence in children: The Binet-Simon Scale*: Williams & Wilkins Company.

Burden, R. (1998). Assessing children's perceptions of themselves as learners and problem-solvers: The construction of the Myself-as-Learner Scale (MALS). *School Psychology International, 19*(4), 291-305.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (1998). The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. *Personality and Individual Differences, 26*(1), 129-140.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29*(6), 1057-1068.

Caprara, G. V., Vecchione, M., Alessandri, G., Gerbino, M., & Barbaranelli, C. (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. *British Journal of Educational Psychology, 81*(1), 78-96.

Cesarini, D., & Visscher, P. M. (2017). Genetics and educational attainment. *npj Science of Learning, 2*(1), 4.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of personality, 17*(3), 237-250.

Conard, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality, 40*(3), 339-346.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13-21.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics, 9*(3), 403-411.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment, 91*(2), 166-174.

Dudbridge, F. (2003). Pedigree disequilibrium tests for multilocus haplotypes. *Genetic Epidemiology, 25*(2), 115-121.

Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2014). PRSice: polygenic risk score software. *Bioinformatics, 31*(9), 1466-1468.

Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual differences, 34*(7), 1225-1243.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research, 74*(1), 59-109.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological science, 15*(6), 373-378.

Fuchsberger, C., Abecasis, G. R., & Hinds, D. A. (2014). minimac2: faster genotype imputation. *Bioinformatics, 31*(5), 782-784.

Geramian, S. M., Mashayekhi, S., & Ninggal, M. T. B. H. (2012). The relationship between personality traits of international students and academic achievement. *Procedia-Social and Behavioral Sciences, 46*, 4374-4379.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry, 38*(5), 581-586.

Gratten, J., Wray, N. R., Keller, M. C., & Visscher, P. M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature Neuroscience, 17*(6), 782-790.

Haworth, C., Davis, O., & Robert, P. (2013). Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics, 16*(1), 117-125.

Heaven, P. C., & Ciarrochi, J. (2008). Parental styles, conscientiousness, and academic performance in high school: A three-wave longitudinal study. *Personality and Social Psychology Bulletin, 34*(4), 451-461.

Hirschberg, N., & Itkin, S. (1978). Graduate student success in psychology. *American Psychologist, 33*(12), 1083-1093.

Holland, D. C., Dollinger, S. J., Holland, C. J., & Macdonald, D. A. (1995). The relationship between psychometric intelligence and the five-factor model of personality in a rehabilitation sample. *Journal of Clinical Psychology, 51*(1), 79-88.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics, 44*(8), 955-959.

Huebner, E. S. (1994). Preliminary development and validation of a multidimensional life satisfaction scale for children. *Psychological Assessment, 6*(2), 149-158.

Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology, 74*(8), 2204-2214.

Jolliffe, I. T. (1986). Principal component analysis and factor analysis *Principal component analysis* (pp. 115-128): Springer.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*(3), 187-200.

Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (1999). The WISC-III as a process instrument. *San Antonio, TX: The Psychological Corporation*.

Kashdan, T. B., Rose, P., & Fincham, F. D. (2004). Curiosity and exploration: Facilitating positive subjective experiences and personal growth opportunities. *Journal of Personality Assessment, 82*(3), 291-305.

Kelly, W. E., & Johnson, J. L. (2005). Time Use Efficiency and the Five-Factor Model of Personality. *Education, 125*(3), 511-514.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior genetics, 23*(1), 21-27.

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2017). *Behavioral Genetics* (7th ed.). New York: Worth Publishers.

Komarraju, M., & Karau, S. J. (2005). The relationship between the big five personality traits and academic motivation. *Personality and Individual Differences, 39*(3), 557-567.

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., . . . Dale, P. S. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences, 111*(42), 15273-15278.

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual differences, 42*(3), 441-451.

Lakes, K. D., Swanson, J. M., & Riggs, M. (2012). The reliability and validity of the English and Spanish strengths and weaknesses of ADHD and normal behavior rating scales in a preschool sample: Continuum measures of hyperactivity and inattention. *Journal of Attention Disorders, 16*(6), 510-516.

Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T.-K., Shirali, M., . . . Fawns-Ritchie, C. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature genetics, 50*(1), 6.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research, 46*(2), 137-155.

Mark, M., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics, 9*(5), 356.

Mason, O., Linney, Y., & Claridge, G. (2005). Short scales for measuring schizotypy. *Schizophrenia Research, 78*(2), 293-296.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities*. New York: Psychological Corporation.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., . . . Sharp, K. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics, 48*(10), 1279-1283.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology, 52*(1), 81-90.

McCrae, R. R., & Costa, P. T. (1997). Conceptions and correlates of openness to experience *Handbook of Personality Psychology* (pp. 825-847): Elsevier.

McCullough, M. E., Emmons, R. A., & Tsang, J.-A. (2002). The grateful disposition: a conceptual and empirical topography. *Journal of personality and social psychology, 82*(1), 112-127.

Mieloo, C., Raat, H., van Oort, F., Bevaart, F., Vogel, I., Donker, M., & Jansen, W. (2012). Validity and reliability of the Strengths and Difficulties Questionnaire in 5–6 year olds: differences by gender or by parental education? *PloS one, 7*(5), e36805.

Mottus, R., Realo, A., Vainik, U., Allik, J., & Esko, T. (2017). Educational attainment and personality are genetically intertwined. *bioRxiv*, 078014.

Mullins-Sweatt, S. N., Jamerson, J. E., Samuel, D. B., Olson, D. R., & Widiger, T. A. (2006). Psychometric properties of an abbreviated instrument of the five-factor model. *Assessment, 13*(2), 119-137.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30-38.

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*(5), 971-990.

Oconnor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*(3), 396-402.

Okbay, A., Baselmans, B. M., De Neve, J.-E., Turley, P., Nivard, M. G., Fontana, M. A., . . . Derringer, J. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics, 48*(6), 624.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Meddens, S. F. W. (2016a). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539.

Organisation for Economic Co-operation and Development. (2007). *Education at a Glance 2007*. Paris, France: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. (2017). *Education at a Glance 2017*. Paris, France: Organisation for Economic Co-operation and Development.

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology, 34*(1), 29-48.

Petrides, K., Chamorro-Premuzic, T., Frederickson, N., & Furnham, A. (2005). Explaining individual differences in scholastic behaviour and achievement. *British Journal of Educational Psychology, 75*(2), 239-255.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular psychiatry, 20*(1), 98.

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*(3), 148-159.

Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics, 47*(7), 702-709.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322-338.

Raven, J., & Raven, J. (1998). Court, JH (1998). *Manual for Raven's progressive matrices and vocabulary scales*.

Raven, J., Raven, J. E., & Court, J. H. (1989). *Mill Hill vocabulary scale*: Psychological Corporation.

Ray, A., & Margaret, W. (2003). *PISA Programme for international student assessment (PISA) PISA 2000 technical report: PISA 2000 technical report*: oecd Publishing.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353-387.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., . . . Agrawal, A. (2013b). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *science, 340*(6139), 1467-1471.

Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., . . . Barendse, M. (2011). The lavaan package for R. Retrieved from http://lavaan.ugent.be/

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of personality and social psychology, 67*(6), 1063-1078.

Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational psychology review, 1*(3), 173-208.

Selzam, S., Dale, P. S., Wagner, R. K., DeFries, J. C., Cederlöf, M., O'Reilly, P. F., . . . Plomin, R. (2017). Genome-wide polygenic scores predict reading performance throughout the school years. *Scientific Studies of Reading, 21*(4), 334-349.

Selzam, S., Krapohl, E., von Stumm, S., O'reilly, P., Rimfeld, K., Kovas, Y., . . . Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular Psychiatry, 22*(2), 267-272.

Sidak, Z. (1971). On probabilities of rectangles in multivariate Student distributions: their dependence on correlations. *The Annals of Mathematical Statistics*, 169-175.

Silverman, W. K., Fleisig, W., Rabian, B., & Peterson, R. A. (1991). Childhood anxiety sensitivity index. *Journal of Clinical Child and Adolescent Psychology, 20*(2), 162-168.

Snyder, C. R., Hoza, B., Pelham, W. E., Rapoff, M., Ware, L., Danovsky, M., . . . Stahl, K. J. (1997). The development and validation of the Children’s Hope Scale. *Journal of pediatric psychology, 22*(3), 399-421.

Socrates, A., Bond, T., Karhunen, V., Auvinen, J., Rietveld, C., Veijola, J., . . . O'Reilly, P. (2017). Polygenic risk scores applied to a single cohort reveal pleiotropy among hundreds of human phenotypes. *bioRxiv*, 203257.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics, 14*(7), 483-495.

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*(4), 363-374.

Swanson, J. M., Schuck, S., Porter, M. M., Carlson, C., Hartman, C. A., Sergeant, J. A., . . . Lakes, K. (2012). Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. *The International journal of educational and psychological assessment, 10*(1), 51.

The International Schizophrenia Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460*(7256), 748-752.

Trautwein, U., & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-students differences and within-student variation. *Journal of educational psychology, 99*(2), 432.

Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of personality and social psychology, 97*(6), 1115-1128.

Trzaskowski, M., Eley, T. C., Davis, O. S., Doherty, S. J., Hanscombe, K. B., Meaburn, E. L., . . . Plomin, R. (2013). First genome-wide association study on anxiety-related behaviours in childhood. *PloS one, 8*(4), e58676.

Van Der Sluis, S., Verhage, M., Posthuma, D., & Dolan, C. V. (2010). Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PloS one, 5*(11), e13929.

Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41*(1), 221-229.

Webb, E. (1915). *Character and intelligence: An attempt at an exact study of character*: University Press.

Wechsler, D., Golombok, S., & Rust, J. (1992). WISC-III UK Wechsler intelligence scale for children: UK manual. *Sidcup, UK: The Psychological Corporation*.

Zeidner, M., & Matthews, G. (2000). *Intelligence and personality*. New York, NY, US: Cambridge University Press.

Chapter 3 – Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them

This chapter, investigating the unique prediction of school type on academic achievement, is presented as a published paper. It is an exact copy of this publication.

Smith-Woolley, E., Pingault, J-B., Selzam, S., Rimfeld, K., Krapohl, E., von Stumm, S., Asbury, K., Dale, P., Young, T., Allen, R., Kovas, Y., & Plomin, R. (2018). Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them. *npj Science of Learning* ***3***, Article no: 3.

Supplementary materials for this chapter, as detailed in the text, are attached as Appendix 2.

npj | Science of Learning www.nature.com/npjscilearn

ARTICLE OPEN

Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them

Emily Smith-Woolley[1], Jean-Baptiste Pingault[1,2], Saskia Selzam[1], Kaili Rimfeld[1], Eva Krapohl[1], Sophie von Stumm[3], Kathryn Asbury[4], Philip S. Dale[5], Toby Young[6], Rebecca Allen[7], Yulia Kovas[8,9] and Robert Plomin[1]

On average, students attending selective schools outperform their non-selective counterparts in national exams. These differences are often attributed to value added by the school, as well as factors schools use to select pupils, including ability, achievement and, in cases where schools charge tuition fees or are located in affluent areas, socioeconomic status. However, the possible role of DNA differences between students of different schools types has not yet been considered. We used a UK-representative sample of 4814 genotyped students to investigate exam performance at age 16 and genetic differences between students in three school types: state-funded, non-selective schools ('non-selective'), state-funded, selective schools ('grammar') and private schools, which are selective ('private'). We created a genome-wide polygenic score (GPS) derived from a genome-wide association study of years of education (*EduYears*). We found substantial mean genetic differences between students of different school types: students in non-selective schools had lower *EduYears* GPS compared to those in grammar ($d = 0.41$) and private schools ($d = 0.37$). Three times as many students in the top *EduYears* GPS decile went to a selective school compared to the bottom decile. These results were mirrored in the exam differences between school types. However, once we controlled for factors involved in pupil selection, there were no significant genetic differences between school types, and the variance in exam scores at age 16 explained by school type dropped from 7% to <1%. These results show that genetic and exam differences between school types are primarily due to the heritable characteristics involved in pupil admission.

npj Science of Learning (2018)3:3 ; doi:10.1038/s41539-018-0019-8

INTRODUCTION

Achievement at the end of full-time compulsory education represents a major tipping point in life, opening up avenues for higher education, including university and beyond. Therefore, understanding the potential predictors of academic achievement at this juncture is of great importance. One such predictor that has been hotly debated is school type. In England, when students transition from primary to secondary school at age 11, they have the option of attending one of three school types. Ninety-three percent of children attend state-funded schools, the majority of which are non-selective[1] (state non-selective). A small proportion of state-funded schools (163 schools out of 3113 schools in England) are academically selective 'grammar' schools. These schools select their intake based on achievement and ability, assessed by an entrance exam. The remainder of students (approximately 7%), are private educated. As well as being fee-paying, private schools are often also academically selective. These school types are assumed to set children on different trajectories, with research linking selective schools (grammar and private schools) to later success, including higher levels of academic achievement, acceptance at university, and even higher earning potential compared to pupils educated in non-selective schools.[2–4]

However, by design, selective schools are able to choose their student intake based on certain pupil characteristics. This can include selection on ability or achievement on an entrance test; both of which have been shown to correlate positively with life outcomes, including later academic achievement.[5,6] Furthermore, by virtue of being fee-paying, entrance into private schools is usually dependent on whether the family can afford it (their socioeconomic status (SES)), which also correlates with future outcomes.[7–10] Even for state schools, family SES may play a role in what school type a student attends, with grammar schools typically located in more affluent areas and attracting higher SES students on average.[11] It is, therefore, possible that improved outcomes for pupils in selective schools do not necessarily reflect a higher quality of education, but may simply be the consequence of selection—either active, as in the case of ability or achievement, or passive, as in the case of family SES.

Given the considerable fees charged by private schools, in addition to the potential stress of selective school entrance exams,

[1]King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, London SE5 8AF, UK; [2]Clinical, Educational & Health Psychology, Division of Psychology & Language Sciences, Faculty of Brain Sciences, University College London, 26 Bedford Way, London WC1H 0DS, UK; [3]London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK; [4]Department of Education, Psychology in Education Research Centre, University of York, York YO10 5DD, UK; [5]Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, NM, USA; [6]New Schools Network, 3 Albert Embankment, London SE1 7SP, UK; [7]Education Datalab, 1st Floor, 11 Tufton Street, London SW1P 3QB, UK; [8]Laboratory for Cognitive Investigations and Behavioural Genetics, Tomsk State University, Lenin Avenue, 36, Tomsk Oblast, Tomsk 634050, Russia and [9]Department of Psychology, Goldsmiths, University of London, 8 Lewisham Way, London SE14 6NW, UK
Correspondence: Emily Smith-Woolley (emily.smith-woolley@kcl.ac.uk)

Received: 18 July 2017 Revised: 17 November 2017 Accepted: 9 February 2018
Published online: 23 March 2018

why do families choose these schools? Among the many possible reasons is superior academic achievement. The finding that pupils at selective schools outperform their non-selective school counterparts in exams has been frequently reported.[2–4,12,13] At age 16, students in the UK typically take the General Certificate of Secondary Education (GCSE) exams. The UK Department for Education shows that 99% of grammar school students obtain top GCSE grades (A*–C grade) in English and mathematics, compared to 64% for all state-funded mainstream school students.[14] However, academic achievement at age 16 is positively correlated with the factors involved in pupil selection, such as prior achievement, ability and SES.[6,15] Therefore, this raises the question—are selective schools adding anything over and above these factors in the prediction of academic achievement?

Several studies have attempted to elucidate the effect of school type on achievement over and above factors on which schools can select (for example,[13,16] for a review, see Coe et al.).[3] However, many of these have not been published in peer-reviewed journals—for example[2,3,17,18]—and we are not aware of a recent peer-reviewed study looking at all three school types: state non-selective, grammar and private schools in the UK. However, the non-peer-reviewed reports support the conclusion that there are only small academic advantages to attending a selective school, after student factors such as achievement, ability and family SES have been taken into account.

Traditionally, the relationship between the factors involved in school admission and later achievement have been thought to operate environmentally. For example, parents with higher SES may invest more time in their children's education[19] and can afford more resources (e.g., more books or private tuition), which in turn may lead to better opportunities and improved achievement. However, a less frequently investigated factor influencing both selection factors, as well as achievement, is genetics. In the example above, parents with higher SES are not only passing on educationally relevant environments, but they are also passing on educationally relevant genes, a concept referred to as gene-environment correlation (rGE).

A vast literature from quantitative genetics has shown that genetic factors explain a substantial amount of variance in selection factors, including ability and achievement.[20–23] Heritability estimates of general cognitive ability (*g*) from twin studies range from around 30% in childhood, to 40–50% in adolescence and approximately 60% in adulthood.[21] Twin studies also show that much of the relationship between selection factors, such as *g*, and later achievement, are substantially influenced by genetics.[22–26] Because twins typically grow up in the same family, the aetiology of traits such as family SES, which do not vary between twins, cannot be estimated in this way. However, heritability can be estimated by genome-wide complex trait analysis (GCTA),[27,28] which uses DNA from unrelated individuals to estimate the proportion of phenotypic variance explained by hundreds of thousands of single-nucleotide polymorphisms (SNPs) genotyped on DNA arrays. This method has also shown that genetics accounts for a significant amount of individual differences in family SES,[29,30] as well as *g* and achievement.[31–33]

School type, like SES, does not tend to vary within twin pairs. However, because GCTA requires large sample sizes, it has so far not been possible to look at the genetic differences between students of different school types. However, powerful genome-wide association (GWA) studies of behavioural traits, which test associations between specific SNPs and traits are starting to make this possible. Although individually these SNPs, identified through GWA studies, are of small effect, by summing their effects together it is possible to create a genetic score for each individual in an independent sample, which explains a substantial proportion of the genetic variation.[34–36] These scores, dubbed 'genome-wide polygenic scores' (GPS) are a game-changer for genetic research and have already proved insightful within the area of educational achievement. For example, a recent study[37] using a GPS derived from a 2016 GWA study of years of education (*EduYears*)[38,39] has shown educational achievement scores at age 16 differ as a function of GPS. There was approximately one standard deviation difference between those in the highest GPS septile and those in the lowest; representing almost a whole school grade difference. Furthermore, while 65% of students in the highest GPS septile went on to university, only 37% in the lowest septile progressed to university-level education.

For the first time, we assess differences in a polygenic score for years of education (*EduYears)* between students from three school types: non-selective, grammar and private schools. We predict that selection involving heritable traits such as achievement, ability and family SES will be reflected in the genetic differences between students of different school types. Furthermore, in line with previous literature, we expect that selection will also create large achievement differences between students attending the three school types, which will reduce substantially once controlling for the selection factors.

RESULTS

Polygenic score differences between school types

Students attending different school types (state non-selective, grammar and private schools) differed genetically, as shown by their mean *EduYears* GPS (see Fig. 1, analysis of variance (ANOVA) details in Table S1). Non-selective state school students had significantly lower *EduYears* GPS scores compared to grammar school students ($t = 4.87$, $p < 0.001$) and private school students ($t = 7.17$, $p < 0.001$). These differences translate to more than a third of a standard deviation difference ($d = 0.41$ and 0.37, respectively). There were no significant mean differences in *EduYears* GPS scores between grammar and private school students ($t = 0.44$, $p = 0.66$). There were also no significant mean differences between state non-selective schools in varying selectivity areas (see Table S2 and Supplementary Fig. S1).

Associations between EduYears GPS and selection factors

EduYears GPS was positively correlated with each of the selection factors (see Supplementary Table S3), explaining 2.1% of the variance in ability, 5.2% in achievement and 6.6% in family SES. *EduYears* GPS was also positively correlated with GCSE, explaining 7.6% of the variance in GCSE scores, similar to previous analysis of these data.[37] Because selective schools actively select for achievement and ability and passively select for SES, all of which correlate with *EduYears* GPS, we tested whether mean differences in EduYears GPS remained once controlling for these factors.

We found that, after accounting for the variance explained by heritable selection factors, there were no significant *EduYears* GPS differences between students of the three school types: state non-selective, grammar and private (see Supplementary Fig. S2 and Supplementary Table S4). Similar results also emerged when we looked at differences between state non-selective schools in varying selectivity areas (see Supplementary Table S5 and Supplementary Fig. S3), showing small differences in *EduYears* between school types.

GCSE differences

Supplementary Table S6 and Fig. 2 show unadjusted average GCSE grades for state non-selective, grammar and private school students, as well as average GCSE score adjusting separately for *EduYears* GPS, family SES, prior ability and prior achievement, and for all variables together. Unadjusted GCSEs between school types mirrored unadjusted *EduYears* GPS results, with large differences between non-selective and selective schools (see 'Unadjusted GCSE' in Fig. 2, details in Supplementary Table S6). Indeed, the

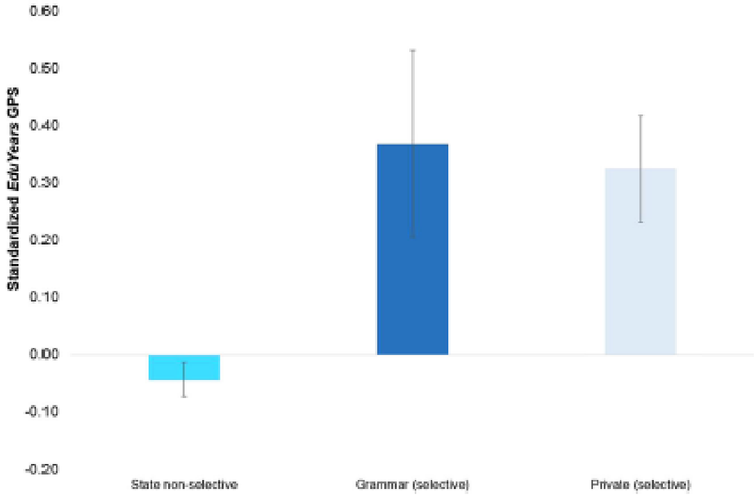


Fig. 1 *EduYears* GPS plotted means (and 95% confidence intervals) between state non-selective, grammar and private school students. *Note:* There were significant *EduYears* GPS mean differences between state non-selective school students and both grammar ($t = 4.869$, $p < 0.001$; $d = 0.413$) and private school students ($t = 7.170$, $p < 0.001$; $d = 0.372$). There was not a significant difference between grammar and private school students ($t = 0.436$, $p = 0.659$)

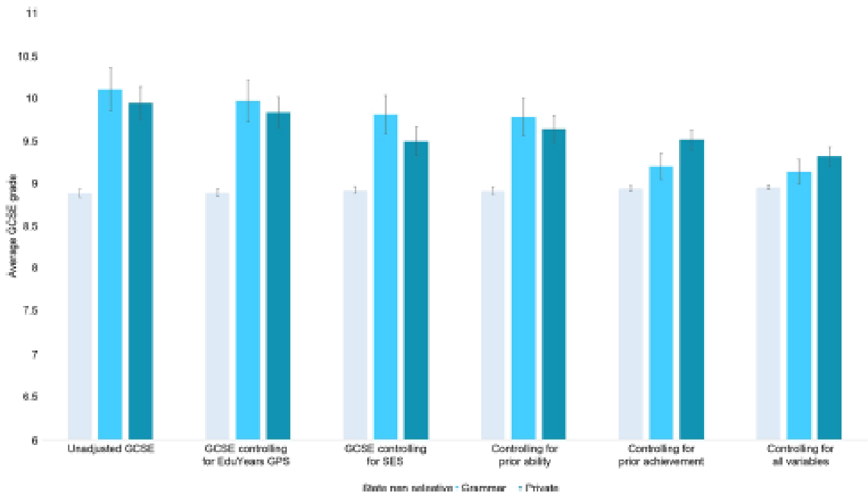


Fig. 2 Plotted means (and 95% confidence intervals) for unadjusted GCSE, GCSE controlling for GPS, GCSE controlling for SES, GCSE controlling for prior ability, GCSE controlling for prior achievement and GCSE controlling for all variables between three school types: state non-selective, grammar and private. *Note:* Details can be found in Supplementary Table S6

mean GCSE score of students attending state non-selective schools was approximately 1 SD below the mean GCSE score of those attending grammar schools ($d = 1.05$, 95% CIs $= 0.83$–1.28) and private school students ($d = 0.92$, 95% CIs $= 0.75$–1.09). This translates to around a whole grade difference between average GCSE scores for state non-selective school students and selective school students. There was no difference between grammar and private school students' average GCSE score ($t = 1.00$, $p = 0.32$). There were also no significant differences between non-selective schools in areas that varied in the selectivity of their schools (see Supplementary Table S7 and Supplementary Fig. S4).

Controlling for selection factors

Controlling for *EduYears* GPS had a small effect on average GCSE grades, with the GCSE variance explained by school type dropping slightly from $R^2 = 0.07$ to 0.06, see Fig. 2, details in Supplementary Table S6). This relatively small effect is to be expected given that *EduYears* GPS accounts for only 8% of the variance in GCSE (see Supplementary Table S3). Controlling for family SES and prior ability had a slightly larger effect on GCSE, in line with the GCSE variance they account for ($R^2 = 24\%$ and 27%, respectively). Out of all of the selection factors, prior achievement had the biggest impact on GCSE grades between school type, with average GCSE for grammar schools falling from 10.12 (grade A) to 9.21 (grade B). After controlling for prior achievement, the variance in GCSE explained by school type dropped from 7.1 to 1.3%.

Controlling for all of the selection factors and *EduYears* GPS together saw a further reduction in average GCSE between school types, with average GCSE score for grammar ($M = 9.14$; $t = 2.35$, $p < 0.019$) and private ($M = 9.32$, $t = 6.16$, $p < 0.001$) similar to that of state non-selective school students' average grade ($M = 8.96$). Although these mean differences between school types remained significant, they were greatly reduced. Standardised betas indicated that attending a grammar school compared to a non-selective state school was associated with an increase of just 0.03 of a standard deviation in GCSE, and for private schools, the increase was 0.07. In addition, no significant differences emerged between non-selective schools in varying selectivity areas (see Supplementary Table S7 and Supplementary Fig. S4).

One of our main findings was that after accounting for the variance explained by the selection factors and *EduYears* GPS, the variance in GCSE explained by school type dropped from 7.1% to only 0.5% (see Supplementary Table S6 for regression results).

DISCUSSION

We report genetic mean differences between students attending three different types of school: state non-selective, grammar and private schools. We find that, on average, students in state non-selective schools have lower polygenic scores for years of education (*EduYears*) compared to their peers in selective schools. Furthermore, following the same pattern of results as *EduYears*, there are also substantial mean differences in GCSE performance between pupils in selective and non-selective school types. However, almost all of these differences are explained by heritable, individual-level factors, which schools actively or passively use in the pupil selection process.

Although finding genetic differences between state non-selective, grammar and private school students may initially seem surprising, when we consider the heritable traits that selection is based on, this difference is less unexpected. Put another way, students with higher polygenic score for years of education have, on average, higher cognitive ability, better grades and come from families with higher SES, and these students are subsequently more likely to be accepted into selective schools. This results in a system in which children are intentionally phenotypically selected, but unintentionally genetically selected.

However, despite finding mean genetic differences between students of different school types, it should be noted that the majority of the variation in *EduYears* GPS occurs within the school type, not between the school types. For example, a Cohen's d of 0.41, (the difference between mean *EduYears* scores for state non-selective school students and grammar school students), which is classed as a small-medium effect size, translates to an overlap of approximately 83% between the two distributions.[40]

Nevertheless, finding an association between genotype and school type suggests that genetic factors are contributing to variation in educational environments, a concept known as gene-environment correlation (rGE). This occurs when individuals select, modify and 'inherit' their environment, in part based on their genotype.[20,41] Putting our research within the context of rGE, we suggest that in addition to students being selected into schools based on their genetically influenced traits (evocative rGE), children themselves also actively select educational environments that correlate with their genotype (active rGE). In the case of high achieving students, these environments might be challenging or competitive academic institutions, which grammar and private schools are often reputed to be. Finally, because we know that the factors used in school selection are substantially heritable, it is likely that academically gifted children will come from academically gifted parents. These parents not only provide the genes but also the environments to help them progress academically.

As well as having a higher average *EduYears* polygenic score, students attending selective schools also achieve better GCSE results on average.[2,3,12–14,17] There has been some debate in the literature as to the size of this achievement gap, with studies accounting for different background characteristics in their analysis. We find that almost all of the selective school advantage in GCSE can be explained by family SES, achievement, ability and *EduYears GPS*. After controlling for these factors, going to a grammar vs. a state non-selective school is associated with a mean GCSE grade increase of just 0.026 of a standard deviation and for private schools, 0.070 of a standard deviation. Furthermore, the variance in GCSE that school type explains falls from 7% to <1%.

Controlling for *EduYears* alone had a fairly small effect on average GCSE grades between school types. However, this is to be expected considering that *EduYears* GPS currently predicts approximately 8% of the variance in GCSE—15% of the heritability estimated by the twin design[22] and approximately one-third of the heritable variance from SNP-based studies of GCSE at age 16.[30] The predictive nature of *EduYears* is likely to increase with more powerful GWA studies. For example, there was a threefold increase in prediction of educational achievement at age 16 from the 2016 *EduYears* GPS (based on a GWA study with $N = 293,723$) as compared to the 2013 *EduYears* GPS ($N = 126,559$).[37]

Although there were only small mean differences between school types once selection factors and *EduYears* were controlled for, this does not mean that other factors are not important for achievement at age 16. Altogether, these factors do not predict all of the variance in GCSE ($R^2 = 0.69$). As shown previously, achievement is the result of many genetically influenced traits, including behaviour, personality, home environment and health.[22] Furthermore, by finding a small effect of school type, we are not saying schools are unimportant, or that teaching does not work. Without schools, it is hard to imagine a successful education system that allows children to reach their academic potential. However, while schools themselves are important for academic achievement, the type of school appears less so. Educational achievement is not necessarily the only reason parents opt to send their children to selective schools. A recent report on private schools found that these students earned about £200,000 more in their early career (between ages 26 and 42) as compared to state school students.[2] However, this report did not distinguish between non-selective and selective state schools. More research is needed to see whether differences in university attendance, career choice and earnings are still predicted by school type once individual student factors have been accounted for. In addition to differences in university and career outcomes, it would also be of interest to identify potential differences between school types in terms of non-cognitive traits as outcomes, with one survey finding 66% of parents believing that private schools 'instil a sense of confidence in pupils'.[2]

There are several limitations to our study. First, we recognise that there is considerable variation in schools within our three school types—within each of the school types, there will be examples of exceptional and under-performing schools. In particular, there is more variance in the state non-selective schools category as it includes most of the schools. It also includes a wide variety of other categories, such as schools that are allowed to select for religion and schools that are allowed to select up to 10% of their pupils for talent in specialist subjects, such as sport, performing or visual arts, and languages. These schools are not allowed to select directly on academic grounds. However, there is some evidence that they do in fact select more able students.[42] Nonetheless, accounting for prior achievement and ability at age 11, before most children enter secondary school, adjusts for this.

Another limitation of the present study is access to school type. Grammar and private schools are not evenly distributed around the country. Therefore, in local authority areas where there are no selective schools, the average GCSE grade of pupils in non-selective schools may be higher and in areas where there are a greater number of selective schools, the average GCSE grade of non-selective schools may be lower. Because there are far fewer selective schools, this geographical effect may potentially inflate the average non-selective school GCSE grade. To see whether this had an impact on GCSE differences, we split the non-selective school group into three further groups: non-selective schools in selective areas, partially selective areas and non-selective areas. Once we controlled for all of the selection factors, we found that there were no differences between non-selective schools in areas of varying selectivity (see Supplementary Table S7 and Supplementary Fig. S4).

A final limitation to note is that the GCSE variable we used in the analysis is a composite of only the three core subjects taken at age 16—English, science and mathematics. For other subjects, such as languages, art and social sciences, school type may have a greater influence. However, because different school types prioritise different subjects,[43] it is difficult to untangle the effect of school type on optional rather than core subjects, although this would be a useful direction for future research.

In the current study, we find genetic differences between students attending three school types: state non-selective schools, grammar schools and private schools. We find that selective school students have higher polygenic scores for years of education on average compared to students attending non-selective schools. Furthermore, we find substantial mean differences in GCSE between school types. However, once student and family factors have been accounted for, as well as *EduYears* GPS, the type of school that a child attends explains less than one percent of the individual differences in educational achievement (GCSE mean grade) at age 16.

METHODS

Sample

This study included unrelated individuals from the Twins Early Development Study (TEDS). TEDS is a large, representative sample of 16,000 twin pairs born in England and Wales between 1994–1996 and followed from birth to the present day.[44] Ethical approval for this study was received from King's College London Ethics Committee. Although there has been some attrition throughout the years, approximately 10,000 twin pairs are still actively involved in the study and provide rich behavioural and cognitive data. Importantly, TEDS was and still is a representative sample of England and Wales, as described in detail elsewhere.[44,45] In the present study, we included 4814 unrelated individuals (one twin randomly in a pair) who had data present for three key variables: genotype data, educational achievement at age 16 and school type data. This sample included 2597 females (54%) and 2217 males (46%). Of this sample, 2533 individuals also had data present for the selection factors: ability, achievement and SES, which included 1427 females (56.3%) and 1106 males (43.7%). For a breakdown of sample sizes by school type, see Supplementary Table S8. Written informed consent was given for all participants involved for each wave of data collection.

Genotyping

For information on how the sample were genotyped and the quality control process, please see Supplementary Methods S1.

Measures

School type. When TEDS twins were 18, they received a questionnaire that included a series of questions asking what type of school they attended when they took exams at age 16—the GCSEs. Respondents were asked to indicate either 'Yes' or 'No' for different school types. We classified all respondents who reported attending either a state non-selective school as 'State non-selective', all those who indicated that they went to a grammar school as 'Grammar' and all those indicating that they went to a private school as 'Private'. In addition to TEDS data, we also accessed school type information through the National Pupil Database (NPD; https://www.gov.uk/government/collections/national-pupil-database). By supplementing TEDS data with that from NPD, our final school type numbers were: state non-selective: $n = 4263$, grammar: $n = 143$, private: $n = 408$. We also further split state non-selective schools into three categories for follow-up analysis: non-selective schools in fully selective areas ($n = 331$), non-selective schools in partially selective areas ($n = 905$) and non-selective schools in non-selective areas ($n = 3027$). For more information on how and why we created these groupings, including accuracy between data sources and selective area groupings, please see Supplementary Methods S2.

Educational achievement at age 16. The GCSE is a standardised UK-based examination administered at the end of compulsory education at age 16 ($M = 16.31$, $SD = 0.29$). Almost all students take the three core subjects: English, mathematics and science. In addition, students are allowed to choose a range of other subjects such as geography, history and art. These subjects were graded from 4 (G, the minimum pass grade) to 11 (A*, the best possible grade). In the current sample, GCSE results were obtained from questionnaires sent via mail, in addition to telephone interviews with twins and their parents. We further supplemented this with data from NPD. Our analyses focused on the three core subjects: English, mathematics and science. Students taking science GCSE are either awarded separate GCSEs for physics, chemistry and biology ('triple science') or as one course, which is double weighted ('double science'), therefore, we took a mean grade of the science GCSEs. Because English, mathematics and science grades correlated highly ($r = 0.70–0.82$), we created a GCSE composite. There were 3920 individuals for whom we had both self-reported GCSE and NPD data, this composite correlated at $r = 0.99$ between both data sources, which supported the high accuracy of TEDS data.

Selection factors

Socioeconomic status. Family SES was measured by taking the arithmetic mean of five measures: maternal and paternal education (measured on a scale from 1–8, where 1 = no education and 8 = postgraduate qualifications), occupation (indexed by the Standard Occupational Classification (2000) on a scale from 1–9, where 1 = elementary administration and service occupations and 9 = managers, directors and senior officials) and maternal age at birth of first child. All measures were standardised to have a mean of 0 and a SD of 1 and at least three measures were required to calculate the arithmetic mean.

Achievement tests at age 11. We did not have access to selective school entrance exams, however, before children transition to secondary school, they are usually required to take exams, which include English and mathematics tests. In our sample these tests comprise two English tests (reading and writing) and three maths tests (calculator and non-calculator test as well as a mental arithmetic test). Due to the high correlation between maths and English scores ($r = 0.67$), we created a composite of these test scores requiring both to be present.

Ability (general cognitive ability, g*).* To measure general cognitive ability, participants were asked to complete an online battery of cognitive tests administered as part of TEDS testing at age 11. These tests included verbal and non-verbal abilities ($M = 11.2$, $SD = 0.69$). A mean score was derived from four tests, two verbal tests (the Wechsler Intelligence Scale for Children (WISC) Vocabulary Multiple-Choice and the WISC General Knowledge test)[46] and two non-verbal tests (Raven's Progressive Matrices[47] and the WISC Picture Completion task).[48]

Data availability

For information on data availability, please see the Twins Early Development Study data access policy. This can be found at: http://www.teds.ac.uk/research/collaborators-and-data/teds-data-access-policy.

Analyses

Genome-wide polygenic scores. We calculated polygenic scores that were based on the summary statistics of the largest GWA study for years of education ($N = 293{,}723$ individuals).[39] A GPS is calculated by using information from GWA study summary statistics about the strength of association between a genetic variant and a trait, to score individuals' genotypes in independent samples. For each genotype in the independent sample, all trait-associated alleles are counted and multiplied by their effect size (i.e., their strength of association with a trait as reported in GWA summary statistics). The sum of these weighted and counted alleles forms a polygenic score for each individual. We used the software PRSice to create individual GPS. Those SNPs that passed quality control were clumped for linkage disequilibrium by applying an $R^2 = 0.1$ cutoff within a 250-kb window. It is possible to calculate various GPS based on different GWA study significance thresholds for genetic variants, with less stringent p-value thresholds resulting in GPS that include more SNPs. Here, we calculated GPS for seven p-value thresholds (0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5). We report analyses for the p-value threshold of 0.05 in the main text; however, the analyses for the other p-value thesholds are reported in Supplementary Fig. S5. We regressed all GPS on the first ten principal components and used these standardised residuals in our analyses to account for population stratification.

Mean differences. To estimate differences between the three school types: state non-selective, grammar and private schools, we used a one-way ANOVA with planned contrasts. In addition to the three-level school type analysis, we also conducted follow-up analysis looking at differences between state non-selective schools in areas with and without grammar schools: non-selective schools in fully selective areas, non-selective schools in partially selective areas and non-selective schools in non-selective areas. As the sample sizes varied between groups, we used adjusted Cohen's *d* to estimate effect size. This test adjusts the calculation of the pooled standard deviation with weights for the sample sizes.

To test the effect of school type after controlling for selection factors (SES, prior achievement and prior ability) and *EduYears* GPS, we conducted hierarchical linear regression with dummy coding. See Supplementary Methods S3 for further information on analysis.

All methods were performed in accordance with relevant regulations and guidelines.

ACKNOWLEDGEMENTS

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a programme grant to R.P. from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ grant agreement n° 602768 and ERC grant agreement n° 295366. RP is supported by a Medical Research Council Professorship award (G19/2). S.S. is supported by the MRC/IoPPN Excellence Award and by the EU Framework Programme 7 (602768). E.K. and K.R. were supported by a Medical Research Council studentship. S.v.S. is supported by a Jacobs Foundation Research Fellowship award (2017–2019). Y.K. is supported by the Tomsk State University competitiveness improvement programme grant 8.1.09.2017. J.B.P. is a fellow of MQ: Transforming Mental Health (MQ16IP16).

AUTHOR CONTRIBUTIONS

R.P. directs and received funding for the Twins Early Development Study (TEDS). R.P. and E.S.W. conceived the present study. E.S.W. analysed and interpreted the data with advice from all co-authors. R.P. supervised the project and interpreted the data. R.P. and E.S.W. wrote the manuscript with help from all authors (J.B.P., S.S., K.R., E.K., S.v.S., K.A., P.S.D., T.Y., R.A., and Y.K.).

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Science of Learning* website (https://doi.org/10.1038/s41539-018-0019-8).

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Department for Education UK Government. *National Statistics: Schools, Pupils and Their Characteristics. National Tables*, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/552342/SFR20_2016_Main_Text.pdf (2016).
2. Broughton, N., Ezeyi, O., Hupkau, C., Keohane, N. & Shorthouse, R. *Open Access: An Independent Evaluation*, available at: http://www.smf.co.uk/wp-content/uploads/2014/07/Open-Access-an-independent-evaluation-Embargoed-00.01-030714.pdf (2014).
3. Coe, R. et al. *Evidence on the Effects of Selective Educational Systems*, available at: https://www.gov.gg/CHttpHandler.ashx?id=97485&p=0 (2008).
4. Dearden, L., Ferri, J. & Meghir, C. The effect of school quality on educational attainment and wages. *Rev. Econ. Stat.* **84**, 1–20 (2002).
5. Schmidt, F. L. & Hunter, J. General mental ability in the world of work: occupational attainment and job performance. *J. Pers. Soc. Psychol.* **86**, 162 (2004).
6. Deary, I. J., Strand, S., Smith, P. & Fernandes, C. Intelligence and educational achievement. *Intelligence* **35**, 13–21 (2007).
7. White, K. R. The relation between socioeconomic status and academic achievement. *Psychol. Bull.* **91**, 461 (1982).
8. Strenze, T. Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* **35**, 401–426 (2007).
9. Sirin, S. R. Socioeconomic status and academic achievement: A meta-analytic review of research. *Rev. Educ. Res.* **75**, 417–453 (2005).
10. Bradley, R. H. & Corwyn, R. F. Socioeconomic status and child development. *Annu. Rev. Psychol.* **53**, 371–399 (2002).
11. Andrews, J., Hutchinson, J. & Johnes, R. *Grammar Schools and Social Mobility*, (Education Policy Institute, London, 2016). Available at: http://epi.org.uk/wpcontent/uploads/2016/09/Grammar-schools-and-social-mobility_.pdf
12. Bolton, P. *Grammar School Statistics*, available at: http://researchbriefings.files.parliament.uk/documents/SN01398/SN01398.pdf (2017).
13. Sullivan, A. & Heath, A. *State and Private Schools in England and Wales*. (University of Oxford, 2002).
14. Department for Education UK Government. *Revised GCSE and Equivalent Results in England*, available at: https://www.gov.uk/government/statistics/revised-gcse-and-equivalent-results-in-england-2015-to-2016 (2016).
15. Goldstein, H. & Sammons, P. The influence of secondary and junior schools on sixteen year examination performance: A cross-classified multilevel analysis. *Sch. Eff. Sch. Improv.* **8**, 219–230 (1997).
16. Clark, D. Selective schools and academic achievement. *BE J Econ Anal Policy* **10**, 20–24 (2010).
17. Anderson, K., Gong, X., Hong, K. & Zhang, X. Do selective high schools improve student achievement? Effects of exam schools in China. *China Econ. Rev.* **40**, 121–134 (2016).
18. Atkinson, A., Gregg, P. & McConnell, B. *The Result of 11 plus Selection: an Investigation Into Opportunities and Outcomes for Pupils in Selective Leas.* (Centre for Market and Public Organisation Working Paper, UK, Bristol, 2006).
19. Waldfogel, J. & Washbrook, E. *On Your Marks: Measuring the School Readiness of Children in Low-to-middle Income Families* (Resolution Foundation, London, 2011).
20. Knopik, V. S., Neiderheiser, J., DeFries, J. C. & Plomin, R. *Behavioral Genetics* 7th edn (Worth Publishers, New York, 2017).
21. Plomin, R. & Deary, I. J. Genetics and intelligence differences: five special findings. *Mol. Psychiatry* **20**, 98–108 (2015).
22. Krapohl, E. et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc. Natl Acad. Sci.* **111**, 15273–15278 (2014).
23. Wainwright, M. A., Wright, M. J., Luciano, M., Geffen, G. M. & Martin, N. G. Multivariate genetic analysis of academic skills of the Queensland core skills test and IQ highlight the importance of genetic g. *Twin Res. Human. Genet.* **8**, 602–608 (2005).
24. Bartels, M., Rietveld, M. J., Van Baal, G. C. M. & Boomsma, D. I. Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Res.* **5**, 544–553 (2002).
25. Thompson, L. A., Detterman, D. K. & Plomin, R. Associations between cognitive abilities and scholastic achievement: Genetic overlap but environmental differences. *Psychol. Sci.* **2**, 158–165 (1991).
26. Calvin, C. M. et al. Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behav. Genet.* **42**, 699–710 (2012).
27. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. *Genome-wide Association Studies and Genomic Prediction*(GCTA): Methods, Data Analyses, and Interpretations. In: Genome-Wide Association Studies and Genomic Prediction. Methods in Molecular Biology (Methods and Protocols), vol 1019. (eds Gondro C., van der Werf J., Hayes B.) 215–236 (Humana Press, Totowa, NJ, 2013).
28. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Human. Genet.* **88**, 76–82 (2011).
29. Trzaskowski, M. et al. Genetic influence on family socioeconomic status and children's intelligence. *Intelligence* **42**, 83–88 (2014).
30. Krapohl, E. & Plomin, R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol. Psychiatry* **21**, 437 (2016).
31. Trzaskowski, M., Shakeshaft, N. G. & Plomin, R. Intelligence indexes generalist genes for cognitive abilities. *Intelligence* **41**, 560–565 (2013).
32. Trzaskowski, M. et al. DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. *Behav. Genet.* **43**, 267–273 (2013).
33. Plomin, R. et al. Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychol. Sci.* **24**, 562–568 (2013).
34. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
35. Harlaar, N. et al. A behavioural genomic analysis of DNA markers associated with general cognitive ability in 7-year-olds. *J. Child Psychol. Psychiatry* **46**, 1097–1107 (2005).
36. Wray, N. R. et al. Research review: polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* **55**, 1068–1087 (2014).

37. Selzam, S. et al. Predicting educational achievement from DNA. *Mol. Psychiatry* **22**, 267–272 (2017).
38. Rietveld, C. A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
39. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
40. Magnusson, K. *Interpreting Cohen's d Effect Size: An interactive Visualization*, available at: http://rpsychologist.com/d3/cohend/ (2014).
41. Rutter, M. et al. Integrating nature and nurture: Implications of person–environment correlations and interactions for developmental psychopathology. *Dev. Psychopathol.* **9**, 335–364 (1997).
42. Gibbons, S. & Silva, O. Faith primary schools: better schools or better pupils? *J. Labor. Econ.* **29**, 589–635 (2011).
43. Iannelli, C. The role of the school curriculum in social mobility. *Br. J. Sociol. Educ.* **34**, 907–928 (2013).
44. Haworth, C. M., Davis, O. S. & Plomin, R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Res. Human. Genet.* **16**, 117–125 (2013).
45. Kovas, Y. et al. The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monogr. Society Res. Child Dev.* **72**, vii–144 (2007).
46. Kaplan, E., Fein, D., Kramer, J., Delis, D. & Morris, R. *The WISC-III as A Process Instrument* (The Psychological Corporation, New York, 1999).
47. Raven, J. *Raven Progressive Matrices*. In: Handbook of Nonverbal Assessment (ed. McCallum RS.) (Springer, Boston, MA, 2003).
48. Wechsler, D. *Wechsler Intelligence Scale for Children* — Third Edition UK (WISC-IIIUK) Manual (Psychological Corporation, London, 1949).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

Chapter 4 – Ofsted secondary school quality is a poor predictor of student academic achievement and wellbeing

This chapter, estimating the unique prediction of Ofsted school quality ratings on student achievement and wellbeing, has been adapted from a manuscript currently under review at *npj* Science of Learning:

Smith-Woolley, E., Cheeseman, R., Pingault, J-B., von Stumm, S., Asbury, A., Dale, P., Allen, R., Kovas, Y., & Plomin, R. (2018). Ofsted secondary school quality is a poor predictor of student academic achievement and wellbeing. *npj Science of Learning*.

Supplementary materials for this chapter, as detailed in the text, are attached as Appendix 3.

Abstract

In the UK, schools are inspected by an independent government agency, the Office for Standards in Education, Children's Services and Skills (Ofsted). Inspections aim to hold schools to account and promote improvement, with results made available to the public. These reports supposedly index school quality, but to what extent do they predict individual-level outcomes at secondary school, such as achievement, wellbeing or student engagement? The current study employs a UK population-based sample of 4,391 individuals to explore the association between Ofsted-rated secondary school quality and achievement, wellbeing and student engagement at age 16. We found that Ofsted ratings of school quality (rated as 'Inadequate', 'Requires Improvement', 'Good' or 'Outstanding') predicted 4% of the variance in achievement, with those attending schools rated 'Good' (the most common category) achieving a third of a grade better than those in schools that 'Require Improvement'. However, much of this advantage appears to reflect differences between students at school intake, rather than added-value of the school. Indeed, once we accounted for student covariates, namely prior achievement and socioeconomic status, Ofsted ratings explained less than 1% of variance in achievement. Attending a 'Good' school was now associated with only a tenth of a grade boost in achievement compared to those in schools that 'Require Improvement'. Furthermore, Ofsted-rated school quality was a poor predictor of school engagement or student wellbeing, with an average correlation of .03. Taken together, this calls into question the usefulness of Ofsted ratings for parents and students when choosing secondary schools.

Introduction

In the UK, parents can choose where to send their children to secondary school. To help with this decision-making process, many turn to the reports by the Office of Standards in Education, Children's Services and Skills (Ofsted). Ofsted is an independent government agency whose purpose is to "inspect and regulate services that care for children and young people" (Ofsted, 2018). The primary aims of these inspections are to drive improvement within schools and hold them to account. School inspections typically happen once every four years and comprise lesson inspections, teacher meetings, paperwork checks and pupil interviews. Once an inspection has been conducted, a school is awarded an overall effectiveness rating. This score falls into one of four categories: 'Outstanding' (23% of schools receive this rating), 'Good' (56%), 'Requires Improvement' (15%) or 'Inadequate' (6%). For those schools that are deemed to be 'Outstanding', this rating can act as a marketing tool, attracting interest from parents, students, potential teachers (Waterreus, 2003) and even driving up house prices (Black, 1999; Gibbons & Machin, 2008; Leech & Campos, 2003). In contrast, schools that are judged to be underperforming suffer reputational damage and will be placed under further inspection. Although there is no doubt that children should receive an education in a safe and supportive environment, what is less clear is whether attending a school deemed to be high quality by Ofsted is associated with better educational and social-emotional outcomes for children, compared to those attending a school deemed to be 'Inadequate'.

Ofsted inspections

All state-funded schools in England are inspected by Ofsted. In 2017/18, £44 million was spent on 6,079 school inspections, with an average of £7,200 per school inspection (National Audit Office, 2018). The frequency of visits and the length of inspection depends on the school's existing rating. For example, a school judged to be 'Good' at their last inspection will normally receive a one-day short inspection every four years (Ofsted, 2015). At the other end of the rating scale, a school whose overall effectiveness category is judged to be 'Inadequate' will require more regular inspection and can even have their funding agreement terminated (Ofsted, 2015). During Ofsted inspections, inspectors are given a free pass to the school for the time they are there, including: dropping in on lessons, checking students' books and interviewing school leaders, teachers and students. This period can be one of frenzy, teacher exhaustion and stress (de Wolf & Janssens, 2007; Gray & Gardner, 1999) with some teachers reporting feelings of disempowerment and low morale during Ofsted inspections (Hopkins et al., 2016).

After the inspection, schools receive a detailed report which includes the overall effectiveness rating ('Inadequate', 'Requires Improvement', 'Good' or 'Outstanding'). This report is published by Ofsted for each school and available on the internet for anyone to read. In particular, these reports are deemed useful by parents when deciding where to send their children to secondary school. A report of 1,000 parents in the UK (Wespieser, Durbin, & Sims, 2015) found that Ofsted ratings were the third most important factor to parents when choosing a school, after location and suitability to the child's needs. A separate report of over 1,000 parents found a similar result, reporting Ofsted ratings as the second most important source of information for parents choosing schools, after word of mouth from other parents (Ofsted, 2017a).

Ofsted inspections and individual-level outcomes

Why do parents look to Ofsted reports of schools? Because they believe it tells them something about the ethos of the school and the outcomes for the students. But to what extent does the Ofsted rating of a school predict *individual-level* outcomes, such as achievement or student wellbeing? This is presumably what parents and students want to know – does going to a better Ofsted-rated quality school mean better exam results or better wellbeing for their child? We could not find a single study looking at the association between school-level Ofsted ratings and individual-level outcomes.

However, there have been several studies looking at school quality and individual outcomes, measured in other ways. These measures of school quality include student-rated school quality (Keith & Cool, 1992), parent-rated (Gibbons & Silva, 2011), teacher-rated (Hoy, Hannum, & Tschannen-Moran, 1998) and more objective measures of school quality, such as pupil-teacher ratio, percentage of teachers with advanced degrees and pupil expenditure (Eide & Showalter, 1998). These show small to moderate effects of school quality on pupil outcomes. Indeed, a study looking at the effect of school environment on student achievement, as measured by teachers, found a moderate influence of the school environment (standardised betas = .30, p <.01) on maths and English achievement whilst accounting for socioeconomic status (Hoy et al., 1998). However, leadership and teacher quality were only weakly associated with achievement (standardised betas = .11-.19, not all significant).

Accounting for student covariates

Because parental choice of school depends on preferences and resources, students are non-randomly distributed across schools. Furthermore, in some cases, schools use student covariates, such as ability or achievement on tests to select student. Therefore, it is important to account for student covariates in order to look at the unique effect of school quality. One way to do this is to use repeated performance observations (Rivkin, Hanushek, & Kain, 2005) in order to control for fixed factors, such as socioeconomic status or prior achievement. The remaining differences in achievement gains are often thought of as the school's influence on academic progress or 'added value'.

School quality and pupil wellbeing

Besides achievement, parents consider other important factors such as wellbeing, happiness and pupil behaviour when choosing schools (Coldron & Boulton, 1991, 1996). Few studies have looked at the relationship between school quality and pupil wellbeing. One large study using data from the Longitudinal Survey of Young People in England, found that school quality was only weakly associated with pupil happiness and wellbeing at school (Gibbons & Silva, 2011). However, this study used parents' perceptions of school quality, which are potentially biased and not directly comparable across schools. Ofsted, on the other hand, is an independent organisation, and therefore may be more reliable at measuring school quality.

The present research

In the present study, we use a large representative sample of 4,391 individuals for whom we had independent Ofsted quality ratings of their school, as well as extensive information on individual outcomes at age 16. Our primary goal was to investigate whether the overall Ofsted rating was associated with a range of pupil-level outcomes, including academic achievement, wellbeing and school engagement while accounting for differences between students on entry into the school. We predicted significant but weak associations between Ofsted and pupil-level outcomes, which would be reduced when child-specific factors were are considered.

Results

Associations between Ofsted Headline Quality Ratings and achievement

The Ofsted overall quality rating correlated .21 with students' GCSE scores, accounting for 4.4% of the variance. Figure 1 depicts the flow of pupils from the four quality categories to GCSE grades. The figure shows that fewer students in 'Outstanding' schools achieve lower grades as compared to students in schools rated 'Requires Improvement' or 'Inadequate'. Despite the mean differences, what is striking is the variability of GCSE grades obtained by students attending schools of different quality; each school quality category contains students who achieve a wide mix of grades at GCSE. To look at the average GCSE differences between the Ofsted rated categories, we ran ANOVA with polynomial trend analysis and planned contrasts (Table S1). A linear trend best described the relationship between the Ofsted school quality categories (F = 201.96, p = 7.68 x 10^{-45}). Indeed, the difference between 'Inadequate' and 'Requires Improvement' schools was a third of a grade (t = 3.06, p <.05), which was similar to the difference between 'Requires Improvement' and 'Good' (0.30 of a grade; t = 6.35, p <.001), or 'Good' and 'Outstanding' (0.34 of a grade; t – 7.78, p <.001). The biggest GCSE difference was therefore between those attending

‘Inadequate’ schools and those attending ‘Outstanding’ schools, with almost a grade difference (0.94 of a grade; t = 9.93, p <.001). Students attending ‘Inadequate’ schools had a mean GCSE grade of C (M = 8.17, SD = 1.23), whereas those in ‘Outstanding’ schools had a mean GCSE grade of B (M = 9.11, SD = 1.20).

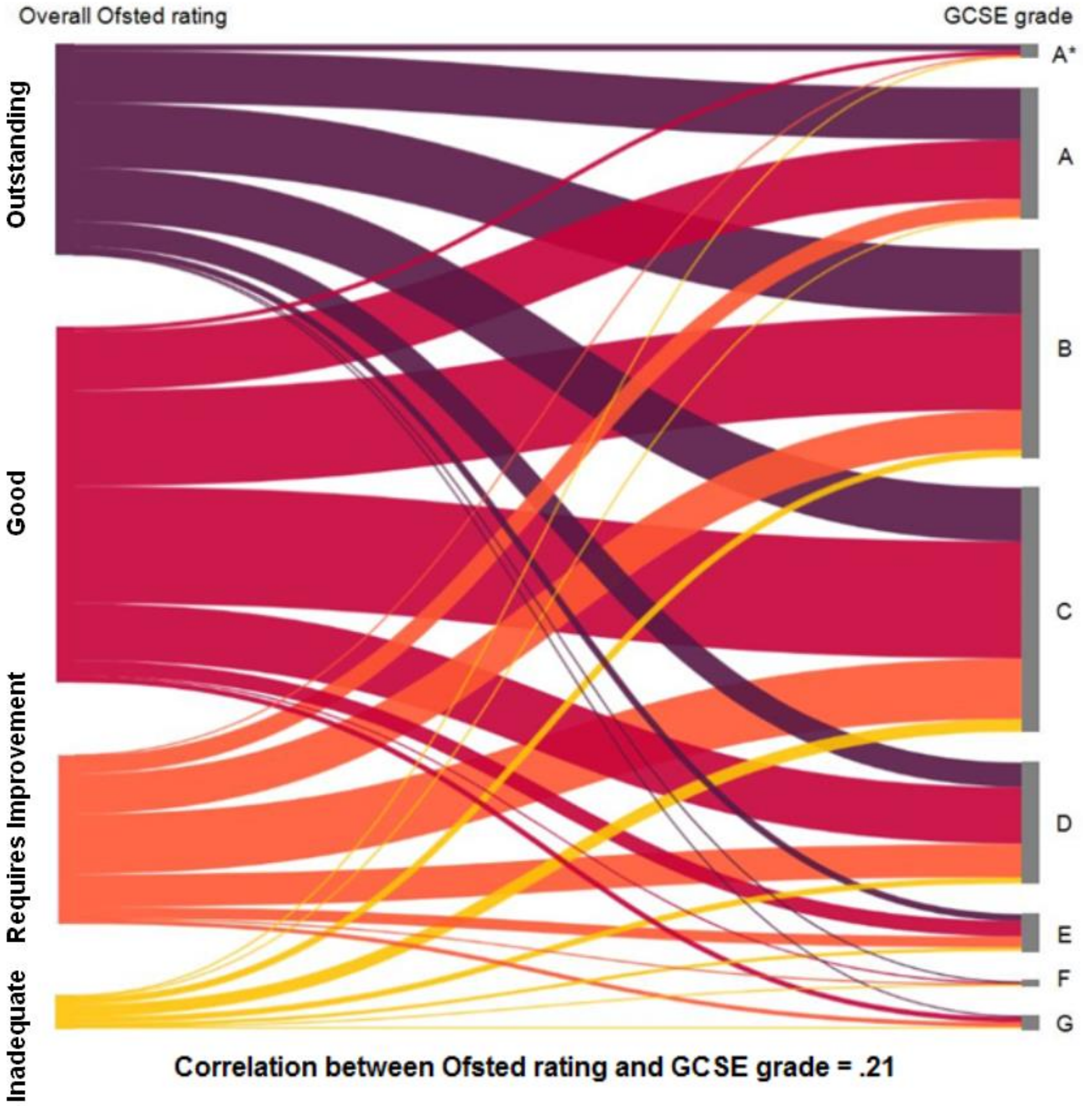


Figure 1. Flow of Ofsted ratings to GCSE grade

Unique prediction of GCSE grades from Ofsted ratings

To explore the unique prediction of the Ofsted overall quality rating on GCSE grades, independent of student covariates, we conducted multiple regression (Table S2). Once we controlled for student covariates, the variance in GCSE predicted by the Ofsted overall quality rating fell from 4.4% to less than 1% (semi-partial correlation = 0.83*)*. Furthermore, the unstandardised beta associated with the Ofsted overall quality rating (*B* = .13) indicated that the average GCSE difference between the categories ('Inadequate'/'Requires Improvement'/'Good'/'Outstanding') was now only a tenth of a grade. To estimate the adjusted means of the Ofsted categories, we ran ANCOVA with pairwise comparisons (Table S3). As previously estimated by the regression, there was roughly a tenth of a grade difference between each of the ordered four categories, with 0.4 of a grade difference at the extremes (between 'Inadequate' and 'Outstanding' schools, $p = 2.91 \times 10^{-9}$). The GCSE difference between attending an Ofsted-rated 'Good' school (the most common Ofsted category) and an 'Outstanding' school is approximately 0.1 of a GCSE grade (p = .001) once student covariates were taken into account. Figure 2 shows the raw and unadjusted GCSE means for each Ofsted school quality category.

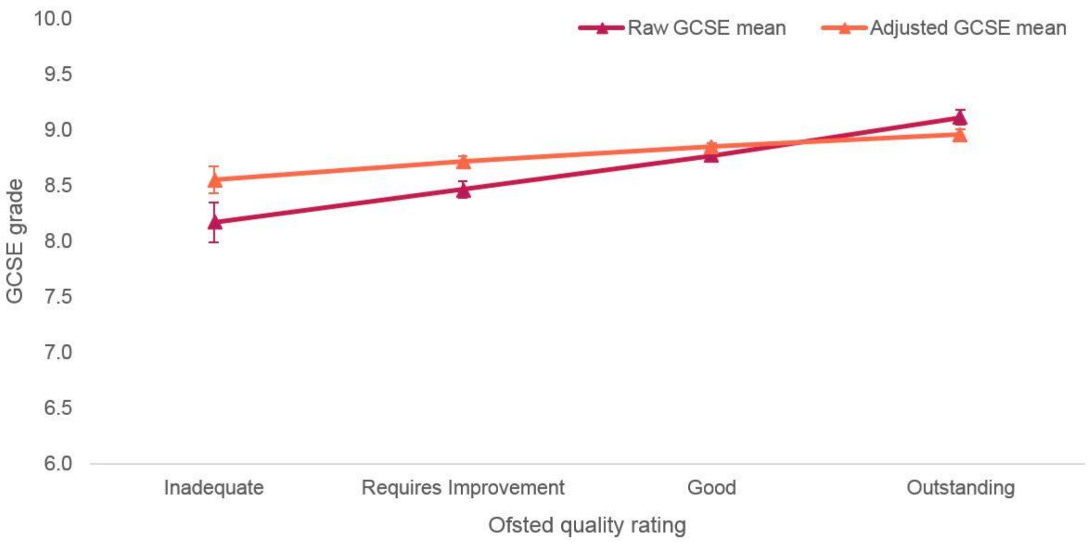


Figure 2. Raw and adjusted GCSE means and 95% confidence intervals. GCSE was graded from 4 (lowest grade: G) to 11 (highest grade: A*).

Associations between Ofsted ratings and students' self-reported experience of the school environment and wellbeing

Finally, we investigated the extent to which Ofsted ratings are associated with student-reported school engagement and wellbeing. The Spearman's correlations between the Ofsted overall quality rating and the 14 student-reported measures of wellbeing and engagement ranged from -.04 (Ambition) to .07 (Homework behaviour), with an average correlation of .03 (see Figure S1). After correction for multiple testing, only the correlation with Homework behaviour remained significant. We ran a further series of ANOVAs which supported these results (see Table S4). Figure 3 shows the means and 95% confidence intervals for wellbeing and school environment measures for students in schools rated as 'Inadequate', 'Requires Improvement', 'Good' and 'Outstanding'. It shows that students attending 'Inadequate' rated schools reported similar levels of happiness, attitudes to school, homework, student teacher relations and ambition as those attending 'Outstanding' rated schools.

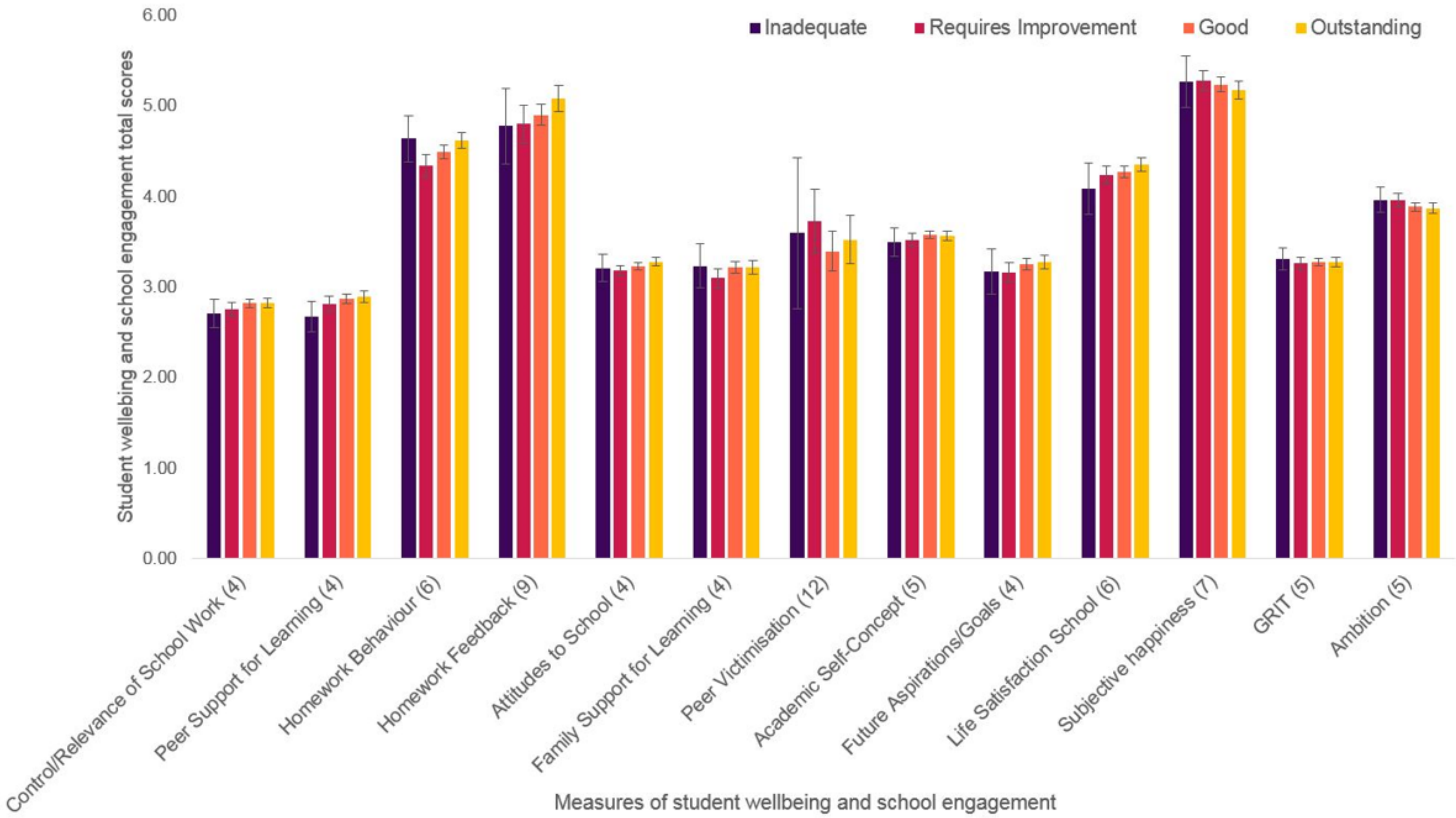


Figure 3. Mean differences (and 95% confidence intervals) for wellbeing and school experiences measures for students attending schools rated as: Inadequate, Requires Improvement, Good and Outstanding by Ofsted. *Note: the total scores for each of the scales are in brackets.*

Discussion

The purpose of this study was to explore the relationship between school quality as rated by Ofsted and outcomes for pupils. We found that the Ofsted overall quality rating predicted 4.4% of the differences in achievement at age 16. However, most of this prediction was accounted for by family socioeconomic status and prior achievement. This indicates that achievement differences between schools of varying quality are largely a result of their intake. In addition, we found that Ofsted-rated school quality was a poor predictor of student wellbeing and student engagement.

Ofsted states that their ratings “allow parents to make informed decisions about where to educate their children” (Ofsted strategy 2017-22, p3). Indeed, one of Ofsted’s priorities is to make their reports “better focused on the issues that parents care about when choosing or seeking assurances about a school” (p9). However, we find that the factors that parents care about (achievement and wellbeing) are only weakly predicted by Ofsted. The correlation between Ofsted ratings of a school and individual-level achievement was .21, explaining 4% of the variance in GCSE grades. However, after accounting for student covariates of socioeconomic status and prior achievement, Ofsted ratings of a school predict less than 1% of the observed differences in examination grades at GCSE. The average GCSE difference between schools of varying quality was just a tenth of a GCSE grade. Put another way, attending a ‘Good’ school over a ‘Requires improvement’ school is associated with a GCSE boost of just 0.1 of a grade, on average.

By statistically accounting for student covariates, such as prior achievement, in the prediction of GCSEs, we generate a proxy of academic progress. Academic progress (referred to as ‘Progress 8’ by the Department for Education), is calculated as achievement at age 16 independent of previous achievement at 11, and is thought to index value added by schools. In other words, academic progress is the difference between examination results at age 16 and what can be predicted by students who achieved the same result at age 11. In the present study, we find that Ofsted-rated quality of a school has little impact on the progress students make during secondary school.

This finding is important for two reasons. Firstly, in a survey of parent views (Ofsted, 2017a) 32% of parents with children aged 0-18 said that they would want to find out about children’s progress in maths at a school when deciding on where to send their child. However, if this is poorly predicted at secondary-school level by Ofsted-rated school quality, they should look to other sources of information when choosing secondary schools. Although we did not calculate effects for individual subjects (such as mathematics or English), given the high correlations between these subjects (r = 0.70–0.82) it is likely that similar results would emerge for individual subjects as we have found for a core subject composite. Secondly, it highlights that the examination differences between students attending different quality schools are largely accounted for by the school intake; better quality schools admit brighter students. This is in line with previous research suggesting that when schools are responsible for their own admissions, they are more likely to select more able pupils (Rivkin et al., 2005; Smith-Woolley et al., 2018; West, 2006).

Although achievement outcomes are important to parents, they are not the only reason why parents opt to send their children to one school over another. In fact, the factors most often cited in parental choice literature are student happiness and wellbeing. In the present study, we find that the correlations between Ofsted ratings and measures of student wellbeing and perceptions of the school environment measures were very small (average r = .03) and non-significant. This suggests that Ofsted-rated school quality has little influence on individual-level wellbeing factors. Put another way, students attending schools with the worst Ofsted ratings report similar levels of happiness, bullying, future aspirations, satisfaction with school and ambition as those students attending schools with the highest Ofsted ratings. These results are in line with previous research looking at the relationship between school quality and wellbeing of students (Gibbons & Silva, 2011). They find that parent-rated school quality is not strongly associated with pupil happiness and wellbeing at the school. This finding has far-reaching implications for parents, who should look to other sources of information on student wellbeing, bullying and student-teacher relations.

There are several limitations to our study. First, we do not consider the impact of school quality at younger ages. The present study focuses on Ofsted reports of secondary schools only. School quality may be more important at younger ages. Indeed, a review of primary school quality on academic achievement across 29 countries (Heyneman & Loxley, 1983) concluded that the quality of primary schools and teacher quality contributed substantially to student achievement, especially in low-income countries. In the present study, we go some way to account for differences between pupils when they enter secondary school by controlling for achievement and socioeconomic status. However, future research may take a longitudinal approach, looking at the academic trajectories of students moving through varying quality schools.

Another limitation of the present study is the lack of objective measures of student wellbeing and student engagement. In the present study, we use 14 self-report measures; however, students are only able to base their judgements on what they know. It is possible that particular students would be happier at different schools yet, because they only have experience of attending their own school, they are not able to make comparisons. One way to explore this possibility would be to look at students who have attended multiple schools of varying quality and compare their wellbeing and satisfaction levels at each school. However, these students may not be representative of the student population and are often moved for a reason, such as family separation, military deployment, exclusion or bullying. Indeed, students who switch schools are, on average, from lower income families and have greater behaviour problems and social interaction difficulties (Gasper, DeLuca, & Estacion, 2012; Sorin & Iloste, 2006).

An alternative and possible way of measuring student happiness and wellbeing is through teacher reports. Teachers have the benefit of interacting with hundreds of pupils and therefore may be more able to make objective judgements. However, this alternative may be problematic for two reasons. Firstly, teachers do not always have strong relationships with students and therefore might not be able to make informed judgements about student wellbeing (child and teacher ratings are only modestly correlated (De Los Reyes & Kazdin, 2005)). Secondly, teacher reports of child wellbeing may be less informative. Parents are likely to be more interested in the child's perception of the schools.

A final limitation to note is that the current sample was drawn from a twin study. Although we only used one twin in a pair for the current study, being a twin might influence the results. However, our sample appears to be largely representative of the general population for achievement (Table S5) and previous research has shown twins to be broadly representative of the general population for health (Andrew et al., 2001), personality (Johnson, Krueger, Bouchard, & McGue, 2002), psychiatric problems (Kendler, Martin, Heath, & Eaves, 1995) and emotional/behavioural problems (Moilanen et al., 1999).

In the current study, we find that Ofsted-rated school quality is a poor predictor of secondary-school outcomes at age 16, including achievement, wellbeing and student engagement, once student characteristics have been taken into account. These findings have important implications for parents looking to Ofsted results as a source of information about student-level outcomes.

Method

Sample

The sample for this study was drawn from the Twins Early Development Study (TEDS). TEDS is a large, population-based sample of twin pairs born in England and Wales between 1994–1996 and followed from birth to the present day (Haworth, Davis, & Plomin, 2013). Ethical approval for this study was received from King's College London Ethics Committee. In the present study, we included 4,391 unrelated individuals (one twin randomly from a pair) attending 2,209 secondary schools for whom we had Ofsted school quality ratings (there were approximately 2 students per school). Participants with severe medical or psychiatric problems or whose mothers had severe medical complications during pregnancy were excluded from the analysis. We also excluded those who attended non-mainstream schools such as special schools for those with learning disabilities. This sample included 2,403 females (55%) and 1,988 males (45%). Informed consent was given by the parents of all participants involved. This sample of 4,391 individuals is broadly representative of the UK population for a number of education and socioeconomic characteristics (see Table S5).

Measures

Ofsted-rated school quality

Headline quality rating

In the current study, there were 4,391 participants for whom we had the overall Ofsted ratings of their school ('Overall effectiveness: How good is the school?'). Of these, 4% attended a school rated as 'Inadequate', 22% attended a 'Requires Improvement' school, 46% attended a 'Good' school and 27% attended an 'Outstanding' school. These statistics were similar to the national percentages previously reported (Ofsted, 2017b). Ofsted reports are publicly available on the internet for all state-funded secondary schools: https://reports.beta.ofsted.gov.uk/. Few studies have published on the reliability of Ofsted ratings. However in 2015/16, Ofsted carried out inspections on the same schools by different inspectors. Of the 24 schools inspected, inspectors agreed on the outcome in 22 cases (National Audit Office, 2018).For further information on school inspections, please see the school inspections handbook: https://www.gov.uk/government/publications/school-inspection-handbook-from-september-2015.

Individual items

In addition to the Headline quality rating, we also had data available on up to 23 individual inspection items, such as “The extent to which pupils contribute to the school and wider community” and “The schools capacity for sustained improvement”. The number of items we had depended on the length of the Ofsted inspection and the risk criteria addressed in their visit. The inter-correlations among the 23 individual Ofsted items revealed moderate to high associations, with an average correlation of r = .59 (see Figure S2A). See Table S6 for the individual items, along with their sample sizes, means and standard deviations.

To guide our decision on the most appropriate measure of Ofsted-rated school quality to use, we conducted principal components analysis (PCA) on the 23 individual items (Table S7). The scree plot (Figure S3) and item loadings (Table S8) supported one general ‘school quality’ principal component, explaining 59% of the variance. The extracted unrotated component correlated highly with all 23 individual items (Figure 2B; average r = .77;), as well as with the Ofsted overall quality rating (Figure 2B; r = .93). This suggests that the Ofsted overall quality rating captures what is in common among the individual items. This result justified our use of the overall quality rating in subsequent analyses in order to maximise sample size (N of overall quality rating = 4,391; N of Ofsted extracted component, which requires complete data for all items = 1,114).

<u>Outcomes at age 16</u>

Achievement

At the end of compulsory education, students in the UK sit the ‘General Certificate for Secondary Education’ (GCSE) examinations. Almost all students take the three core subjects: English, mathematics and science. In addition, students take a range of other subjects such as geography, history and art. All subjects are coded from 4 (G, the lowest grade) to 11 (A*, the best possible grade). In the current sample, GCSE results were obtained from questionnaires sent via mail, in addition to telephone interviews with twins and their parents when they were 16 (M = 16.6, SD = 0.32). We further supplemented this with data from the National Pupil Database (NPD; https://www.gov.uk/government/collections/national-pupil-database). The NPD is a pupil-level database that matches pupil and school characteristic data to pupil level attainment in England. GCSE scores from NPD and TED correlate at .99, therefore we feel confident to take NPD ratings when TEDS data was missing. There were 4,379 students who had GCSE data and Ofsted data.

In the present study, we focused on the three core subjects: English, mathematics and science, which are taken by all students. Because English, mathematics and science grades correlated highly ($r = 0.70–0.82$), we created a GCSE composite requiring at least two grades to be present.

Student-reported school engagement

At age 16, students answered eight questionnaires about their experience of school engagement, including: teacher-student relations, control over and relevance of school work, peer support for learning, family support for learning, homework behaviour, homework feedback, attitudes to school and peer victimisation. Details about these questionnaires can be found in the Supplementary Measures section of the Supplementary Materials.

Academic wellbeing

At age 16, students also answered six questionnaires relating to their academic wellbeing. These questionnaires assessed: academic self-concept, future aspirations and goals, life satisfaction in relation to school, subjective happiness, grit and ambition. Details about these questionnaires can be found in the Supplementary Measures.

Student covariates

To estimate the relationship between school quality and pupil outcomes more rigorously, we considered individual characteristics of students as covariates. We selected two covariates that previous studies have shown to be influential on student achievement: socioeconomic status and prior achievement (Hemmings, Grootenboer, & Kay, 2011; Rimfeld et al., in press; Sirin, 2005).

Socioeconomic status

A measure of family socio-economic status was created by calculating the mean of five measures: maternal and paternal education (measured on a scale from 1–8, where 1 = no education and 8 = postgraduate qualifications), maternal and paternal occupation (indexed by the Standard Occupational Classification on a scale from 1–9, where 1 = elementary administration and service occupations and 9 = managers, directors and senior officials) and maternal age at birth of first child. These measures were collected at first contact, when the sample were 2 years old. All measures were standardised to have a mean of 0 and a SD of 1 and at least three measures were required to calculate the arithmetic mean.

Prior achievement

At age 11, before children transition to secondary school, they are usually required to take exams, which include English, mathematics and science tests. We used the 'fine point score' of each of these tests from the National Pupil Database: https://nationalpupildatabase.wikispaces.com/KS2.

Data availability

For information on data availability, please see the Twins Early Development Study data access policy. This can be found at: http://www.teds.ac.uk/research/collaborators-and-data/teds-data-access-policy.

Analyses:

Associations between Ofsted ratings and individual outcomes

We calculated Spearman's Rank correlation to explore the relationship between the Ofsted overall quality rating and achievement, wellbeing and student engagement measures. In addition to investigating individual differences in outcomes, we also estimated the average differences of students attending schools of different quality using ANOVA with polynomial trend analysis and planned contrasts. Trend analysis tests the relationship between the group means ('Inadequate'/'Requires Improvement'/'Good'/'Outstanding') comparing linear, quadratic and cubic trends. A linear trend would suggest a proportionate change in the value of the outcome across ordered categories, for example GCSE scores increasing proportionately across each Ofsted category ('Inadequate'/'Requires Improvement'/'Good'/'Outstanding'). Quadratic and cubic trends would suggest that the relationship between outcome measures (achievement, wellbeing and student engagement) and Ofsted-rated school quality change across the ordered categories of Ofsted school quality.

To test the effect of Ofsted-rated quality on achievement, independent of student covariates (family socioeconomic status and prior achievement), we conducted regression and observed the unique variance explained by Ofsted-rated school quality. We also investigated the unstandardised beta to get an estimate of the average GCSE difference between different Ofsted-rated schools. Finally, we ran ANCOVA to estimate the adjusted means of the Ofsted-rated school quality categories.

All methods were performed in accordance with relevant regulations and guidelines.

Acknowledgements

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938) and the European Commission (602768; 295366). RP is supported by a Medical Research Council Professorship award (G19/2) and a European Research Council Advanced Investigator award [295366]. RC is supported by an ESRC studentship. JBP is a fellow of MQ: Transforming Mental Health (MQ16IP16). SvS is supported by a Jacobs Foundation Research Fellowship award (2017-2019). YK is supported by the Tomsk State University competitiveness improvement programme grant 8.1.09.2017.

Author contributions:

RP directs and received funding for the Twins Early Development Study (TEDS). RP and ESW conceived the present study. ESW analysed and interpreted the data with advice from all co-authors. RP supervised the project and interpreted the data. RP and ESW wrote the manuscript with help from all authors (RC, JBP, SvS, KA, PSD, RA and YK).

Competing financial interests

The authors declare no conflict of interest.

References

Andrew, T., Hart, D. J., Snieder, H., de Lange, M., Spector, T. D., & MacGregor, A. J. (2001). Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Research and Human Genetics, 4*(6), 464-477.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics, 114*(2), 577-599.

Coldron, J., & Boulton, P. (1991). 'Happiness' as a criterion of parents' choice of school. *Journal of Education Policy, 6*(2), 169-178.

Coldron, J., & Boulton, P. (1996). What do Parents Mean when they Talk About 'Discipline'in Relation to their Children's Schools? *British Journal of sociology of education, 17*(1), 53-64.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483-509.

de Wolf, I. F., & Janssens, F. J. G. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education, 33*(3), 379-396. doi: 10.1080/03054980701366207

Eide, E., & Showalter, M. H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics letters, 58*(3), 345-350.

Gasper, J., DeLuca, S., & Estacion, A. (2012). Switching schools: Revisiting the relationship between school mobility and high school dropout. *American Educational Research Journal, 49*(3), 487-519.

Gibbons, S., & Machin, S. (2008). Valuing school quality, better transport, and lower crime: evidence from house prices. *oxford review of Economic Policy, 24*(1), 99-119.

Gibbons, S., & Silva, O. (2011). School quality, child wellbeing and parents' satisfaction. *Economics of Education Review, 30*(2), 312-331.

Gray, C., & Gardner, J. (1999). The impact of school inspections. *Oxford Review of Education, 25*(4), 455-468.

Haworth, C. M., Davis, O. S., & Plomin, R. (2013). Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics, 16*(1), 117-125.

Hemmings, B., Grootenboer, P., & Kay, R. (2011). Predicting mathematics achievement: The influence of prior achievement and attitudes. *International Journal of Science and Mathematics Education, 9*(3), 691-705.

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high-and low-income countries. *American Journal of sociology, 88*(6), 1162-1194.

Hopkins, E., Hendry, H., Garrod, F., McClare, S., Pettit, D., Smith, L., . . . Temple, J. (2016). Teachers' views of the impact of school evaluation and external inspection processes. *Improving Schools, 19*(1), 52-61.

Hoy, W. K., Hannum, J., & Tschannen-Moran, M. (1998). Organizational climate and student achievement: A parsimonious and longitudinal view. *Journal of School Leadership, 8*, 336-359.

Johnson, W., Krueger, R. F., Bouchard, T. J., & McGue, M. (2002). The personalities of twins: Just ordinary folks. *Twin Research and Human Genetics, 5*(2), 125-131.

Keith, T. Z., & Cool, V. A. (1992). Testing models of school learning: Effects of quality of instruction, motivation, academic coursework, and homework on academic achievement. *School Psychology Quarterly, 7*(3), 207-226.

Kendler, K. S., Martin, N. G., Heath, A. C., & Eaves, L. J. (1995). Self-report psychiatric symptoms in twins and their nontwin relatives: Are twins different? *American Journal of Medical Genetics, 60*(6), 588-591.

Leech, D., & Campos, E. (2003). Is comprehensive education really free? A case study of the effects of secondary school admissions policies on house prices in one local area. *Journal of the Royal Statistical Society, 166*(1), 135-154.

Moilanen, I., Linna, S.-L., Ebeling, H., Kumpulainen, K., Tamminen, T., Piha, J., & Almqvist, F. (1999). Are twins' behavioural/emotional problems different from singletons'? *European child & adolescent psychiatry, 8*(4), S62.

National Audit Office. (2018). Ofsted's inspection of schools. London, UK.

Ofsted. (2015). School Inspection Handbook: Handbook for Inspecting Schools in England Under Section 5 of the Education Act 2005: Ofsted Manchester.

Ofsted. (2017a). *Annual parents survey 2017: parents' awareness and perceptions of Ofsted*. London: Office for Standards in Education, Children's Services and Skills.

Ofsted. (2017b). *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2016/17*. London, UK: Office of Standards in Education.

Rimfeld, K., Malanchini, M., Krapohl, E., Hannighan, L., Dale, P., & Plomin, R. (in press). The stability of educational achievement across school years is largely explained by genetic factors. *npj Sceince of Learning*.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research, 75*(3), 417-453.

Smith-Woolley, E., Pingault, J.-B., Selzam, S., Rimfeld, K., Krapohl, E., von Stumm, S., . . . Allen, R. (2018). Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them. *npj Science of Learning, 3*(1), 3.

Sorin, R., & Iloste, R. (2006). Moving schools: Antecedents, impact on students and interventions. *Australian Journal of Education, 50*(3), 227-241.

Waterreus, J. M. (2003). *Lessons in teacher pay: Studies on incentives and the labor market for teachers.* University of Amsterdam, Netherlands.

Wespieser, K., Durbin, B., & Sims, S. (2015). *School Choice: The Parent View*. Slough: National Foundation for Educational Research.

West, A. (2006). School choice, equity and social justice: the case for more control. *British Journal of Educational Studies, 54*(1), 15-33.

Chapter 5 – The genetics of university success

This chapter, taking a multi-method approach to investigating the genetic influence on university access and success, has been adapted from a manuscript accepted at Scientific Reports:

Smith-Woolley, E[1]., Ayorech, Z[1]., Dale, S., von Stumm, S., & Plomin, R. (2018). The genetics of university success. *Scientific Reports*.

[1] Joint first authors

Supplementary materials for this chapter, as detailed in the text, are attached as Appendix 4.

Abstract

University success, which includes enrolment in and achievement at university, as well as quality of the university, have all been linked to later earnings, health and wellbeing. However, little is known about the causes and correlates of differences in university-level outcomes. Capitalising on both quantitative and molecular genetic data, we perform the first genetically sensitive investigation of university success with a UK-representative sample of 3,000 genotyped individuals and 3,000 twin pairs. Twin analyses indicate substantial additive genetic influence on university entrance exam achievement (57%), university enrolment (51%), university quality (57%) and university achievement (46%). We find that environmental effects tend to be non-shared, although the shared environment is substantial for university enrolment. Furthermore, using multivariate twin analysis, we show moderate to high genetic correlations between university success variables (27-76%). Analyses using DNA alone also support genetic influence on university success. Indeed, a genome-wide polygenic score, derived from a 2016 genome-wide association study of years of education, predicts up to 5% of the variance in each university success variable. These findings suggest young adults select and modify their educational experiences in part based on their genetic propensities and highlight the potential for DNA-based predictions of real world outcomes, which will continue to increase in predictive power.

Introduction

The difference in earnings between high school and university graduates is estimated at $1 million over the course of the lifetime (Carnevale, Cheah, & Hanson, 2015). However, the difference in earnings varies by the type of university attended (Hoekstra, 2009), as well as achievement at university (Jones & Jackson, 1990). Furthermore, the benefits associated with obtaining a university education extend beyond earnings, to include better health and wellbeing, higher rates of employment and even increased life expectancy (Bradley & Corwyn, 2002). Despite this, little is known about the causes and correlates of differences in university-level outcomes, including entrance into university, achievement at university and the quality of university attended.

Differences in who obtains a university degree and who does not are, at least in part, associated with differences in prior academic achievement. A large literature of quantitative genetic studies shows that achievement in childhood and adolescence are substantially heritable, with 40 to 60% of the individual differences in achievement due to genetic factors (Baker, Treloar, Reynolds, Heath, & Martin, 1996; Bartels, Rietveld, Van Baal, & Boomsma, 2002; Branigan, McCallum, & Freese, 2013; Kovas, Haworth, Dale, & Plomin, 2007). However, there are few studies looking at the heritability of academic achievement beyond compulsory education. One twin study, using the same data as in the present study, investigated the heritability of university entrance exams taken at age 18. This study found that the decision to take these exams and students' average grade were both substantially influenced by genetic factors. Estimates ranged from 50% in the humanities to 60% in science, technology, engineering and mathematics (STEM) subjects (Rimfeld, Ayorech, Dale, Kovas, & Plomin, 2016). Interestingly, this study found that, although the influence of the shared environment was minimal for average exam grade, it was substantial for the decision to take these exams or not. The shared environment explained almost half of the liability. This suggests that components of the shared environment, such as family or school, push both members of a twin pair to make the same decisions regarding whether or not to take university entrance exams, but that individual-specific environments influence achievement in university entrance exams. However, the role of shared environmental influence in educational choices at university-level has not yet been studied.

Achievement in high school is highly heritable and stable across development (Rimfeld et al., in press). Indeed, in a longitudinal study using the same sample as in the present study (Rimfeld et al., in press), the heritability of achievement was substantial from age 7 (69%) up to age 16 (61%). In addition, the stability was shown to be driven by genetic effects. For example, 72% of the correlation (r=.66) between age 7 achievement and achievement at age 16 was shown to be due to additive genetic influences. However, the pattern of stability into university remains unclear. Unlike high school, where there is often a relatively uniform curriculum to follow, university provides students with greater opportunity to carve out their interests and choose environments based on their natural abilities and aptitudes. Because these traits are genetically influenced, the university environments that individuals choose might correlate with their genotype. For example, someone who is naturally talented at maths might apply to and attend a university specialising in maths, take extra maths classes or join a maths club. In this way, they have selected environments that correlate with their genetically influenced abilities—a concept known as *gene-environment correlation* (Plomin, 1994; Plomin, DeFries, & Loehlin, 1977). Gene-environment correlation has been shown for traits long assumed to be environmental, including life events (Bolinskey, Neale, Jacobson, Prescott, & Kendler, 2004; Saudino, Pedersen, Lichtenstein, McClearn, & Plomin, 1997), media use (Ayorech, von Stumm, Haworth, Davis, & Plomin, 2017), and occupational status (Fulker & Eysenck, 1979) (for reviews see (Jaffee & Price, 2007; Plomin & Bergeman, 1991). For this reason, choosing to enrol at university, as well as the quality of the institution, are also likely to show genetic influence.

University quality has been assessed using several different indicators, such as academic reputation, employment prospects, research quality and teaching (Brooks, 2005; Ramsden, 1991; Tam, 2001). For example the ‘Complete University Guide’ (https://www.thecompleteuniversityguide.co.uk/league-tables/rankings) takes into account entry standards, student satisfaction, research quality and graduate prospects when ranking UK universities. However, attendance at different quality universities is not random; because the best quality universities can be highly selective (Dill & Soo, 2005), entrance into the universities is therefore at least partly dependent on university entrance exam achievement. Therefore, in order to explore the aetiology of individual differences in the quality of university students attend it will be important to consider its relationship with prior achievement.

Recent advances in molecular genetics have confirmed a genetic contribution to variance in education-related traits. Genome-wide polygenic scores (GPS), which aggregate the effects of thousands of DNA variants identified through genome-wide association (GWA) studies, can be used to predict educational attainment and achievement. One such GPS which has been found to be predictive of many educationally-relevant traits is a GPS for years of education (*EduYears)* (Okbay et al., 2016; Rietveld et al., 2013)). The GWA meta-analysis from which this GPS is derived (Okbay et al., 2016) focused on years of education in a sample of over 300,000 individuals. For example, high school completion counted as 11 years of education, whereas completing a PhD was approximately 20 years. The summary statistics can be applied to an independent sample to create a GPS. Individuals with a high *EduYears* GPS will have many of the genetic variants associated with more years of education, whereas those with a low *EduYears* GPS score will have fewer of these genetic variants. Previously we have shown that *EduYears* GPS explains 2.8% of the variance in academic achievement at age 7, 4.6% at age 12 and 9.1% at age 16 in the current sample (Selzam et al., 2017). In addition to achievement, *EduYears* GPS has also been used to explore ‘environments’, for example *EduYears* GPS predicts 7% of the likelihood of going to university compared to becoming NEET (not in education, employment or training) at age 18 (Ayorech, Plomin, & von Stumm, in press), social mobility (Belsky et al., 2018) and whether students attend selective or non-selective high schools (Smith-Woolley et al., 2018).

In the current study, we use a multi-method approach to investigate the genetics of university success. Capitalising on both twin and molecular genetic data, we perform the first genetically sensitive study of university success, including: achievement in university entrance exams, enrolment at university, university quality (the ranking of the university in league tables), university quality regressed for prior achievement and achievement at university (final degree grade for the whole sample and separately for STEM and humanities subjects). We also explore the genetic links between these variables using multivariate twin analysis.

Results

Phenotypic analyses

Table S1 in the Supplementary Materials lists the sample sizes, means and standard deviations for entrance exam achievement, university quality, university achievement, university enrolment and university quality regressed for university entrance exam achievement, separately for males and females and for zygosity groups. Analysis of variance (ANOVA) was performed on each of the continuous university success variables in order to assess the mean effects of sex, zygosity and their interaction. It can be seen from Table S1 that, although there were mean differences between males and females in entrance exam achievement, university quality and university quality regressed for prior achievement, cumulatively they explain less than 1% of the variance. As a result, sex-limitation model fitting was not performed on these variables and for all subsequent analyses, the data were age and sex regressed and van der Waerden transformed (Van Der Waerden, 1975). Residuals were retained for twin and genomic comparisons. All twin analyses were conducted on the full sample, combining DZ opposite sex and same sex twin pairs.

Sensitivity analyses

Sensitivity analyses comparing entrance exam achievement and university quality between those individuals who did or did not report their final university degree grade were performed. Results indicated that group membership (whether or not a final degree grade was reported) accounted for less than 1% of the variance in entrance exam achievement and university quality suggesting our results were not inflated by missing data (Table S2).

Intraclass twin correlations

Twin correlations for the university success variables can be found in Table S3 in Supplementary Materials. For all of the measures, MZ twin correlations exceeded those of the DZ twins, suggesting genetic influence. For example the MZ correlations for university entrance exam achievement were approximately 0.70, compared to 0.40 for DZ twins. Rough estimates of additive genetic (A), shared environmental influence

(C) and non-shared environmental influence (E) using Falconer's formula (Falconer, 1960) can be found in Table S3.

Twin analysis

To investigate the extent to which genetic and environmental factors explained the variance in each of the university success variables, as well as the associations between them, univariate and multivariate genetic analyses were conducted with OpenMx in the R statistical modelling package (Boker et al., 2011).

Univariate genetic analysis

Univariate twin analyses were performed on the university success variables using structural equation modelling. Here, phenotypic variance in a trait is decomposed into additive genetic (A), shared environmental (C) and non-shared environmental (E) influences. For further information on the univariate twin analysis, see the Method section. We tested full ACE models, as well as nested models (AE and CE) and found that, for university entrance exams and university quality, ACE models fit the data best (Tables S4 and S5). For university achievement and university quality regressed for university entrance exams AE models fit the data best (Tables S6 and S7). For university enrolment, we fit a liability threshold model (Table S8), which decomposes the liability of going to university or not into A, C and E.

Figure 1 shows the variance in each of the university success markers that can be attributed to A, C and E. Table S9 gives these estimates along with their 95% confidence intervals. All of the measures were substantially genetically influenced, with additive genetics accounting for between 46% and 57% of the variance. These estimates were in line with the intraclass twin correlations (Table S3). The heritability of achievement decreases from entrance exam achievement (57%) to university achievement (46%), as did the influence of the shared environment, which was no longer significant for university achievement. University enrolment showed the most shared environmental influence (36%). University quality was highly heritable (57%), with most of the remaining variance explained by the non-shared environment.

Interestingly, the measure of university quality continued to be substantially heritable (47%) even after we accounted for entrance exam achievement. This suggests that the high heritability of university quality reflects more than just prior achievement

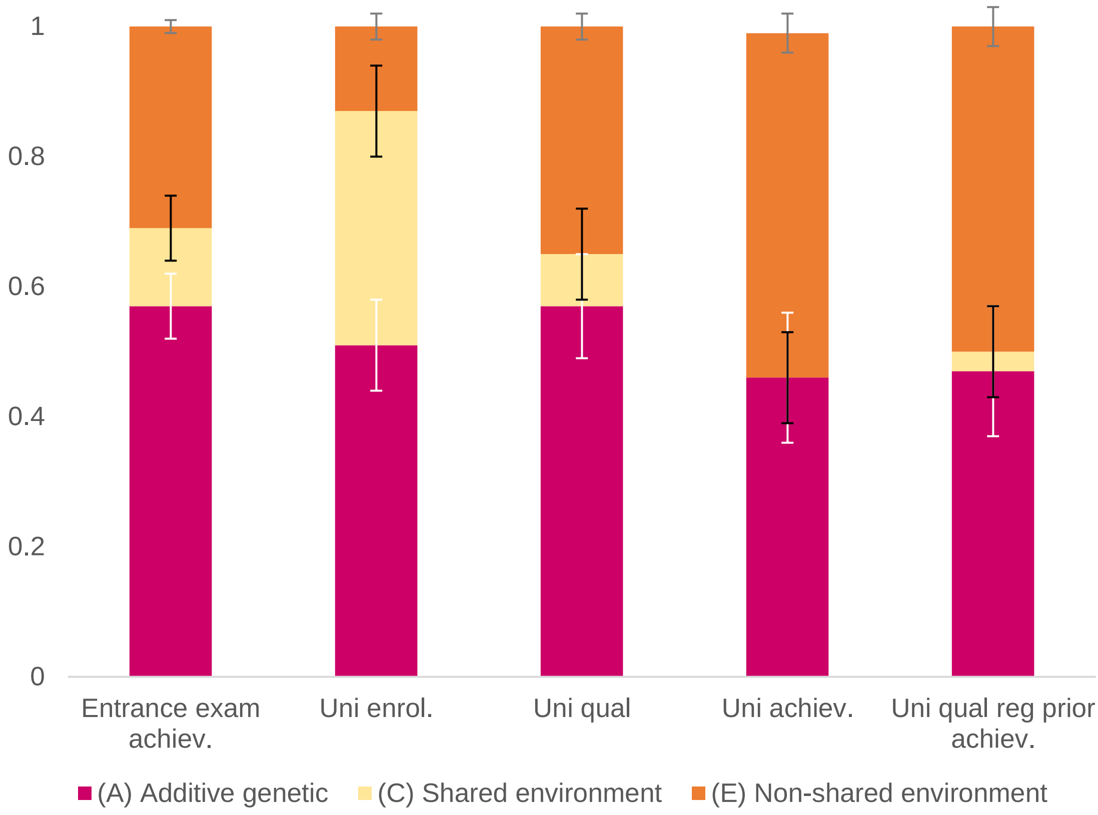


Figure 1 – Model-fitting results and 95% confidence intervals for additive genetic (A), shared environment (C), and non-shared environment (E) components of variance for entrance exam achievement, university enrolment, university quality, university achievement and university quality regressed for entrance exam achievement.

Multivariate genetic analysis

To test the genetic and environmental links between the variables, multivariate genetic analyses were also conducted. To find the best-fitting model, we tested three multivariate designs: correlated factors, common pathway and independent pathway models, and compared their fit statistics (Table S10). Correlated factors was the best-fitting model and the results are presented in this form.

Genetic, shared environmental and non-shared environmental correlations between university success variables can be found in Table S11. As indicated by the genetic correlations (Table S11, part a), there was a high degree of genetic correlation between entrance exam achievement and university quality (r_g= 0.76). Furthermore, there was also a moderate degree of genetic correlation between entrance exam achievement and university achievement (r_g = 0.49). Genetic correlations were weaker between university quality and university achievement (r_g = 0.27). Turning to shared environmental correlations (Table S11, part b), there was a high shared environmental correlation between entrance exam achievement and university quality (r_c = 0.81). Shared environmental correlations between entrance exam achievement and university achievement and between university quality and university achievement were weaker (r_c= 0.35 and 0.27 respectively). Non-shared environmental correlations (Table S11, part c) were moderate between entrance exam achievement and university quality (r_e=0.35), however they were mainly non-overlapping for entrance exam achievement and university achievement (r_e=0.03) and between university quality and university achievement (r_e=0.09).

Polygenic score analysis

To investigate the extent to which SNPs associated with years of education (*EduYears*) predicted the university success variables, we created a genome-wide polygenic score (GPS) and correlated it with our variables (for more information on how this GPS was created, see the Method section). The *EduYears* GPS significantly predicted university success variables, explaining 4% of the variance in entrance exam achievement, 5% in university enrolment, 2% in university quality and 0.7% in university achievement (Figure 2, for values, see Table S12). Furthermore, there was no difference in prediction of *EduYears* GPS between achievement in humanities subjects compared to achievement in STEM subjects (z = 1.08, p = .28; see Table S13).

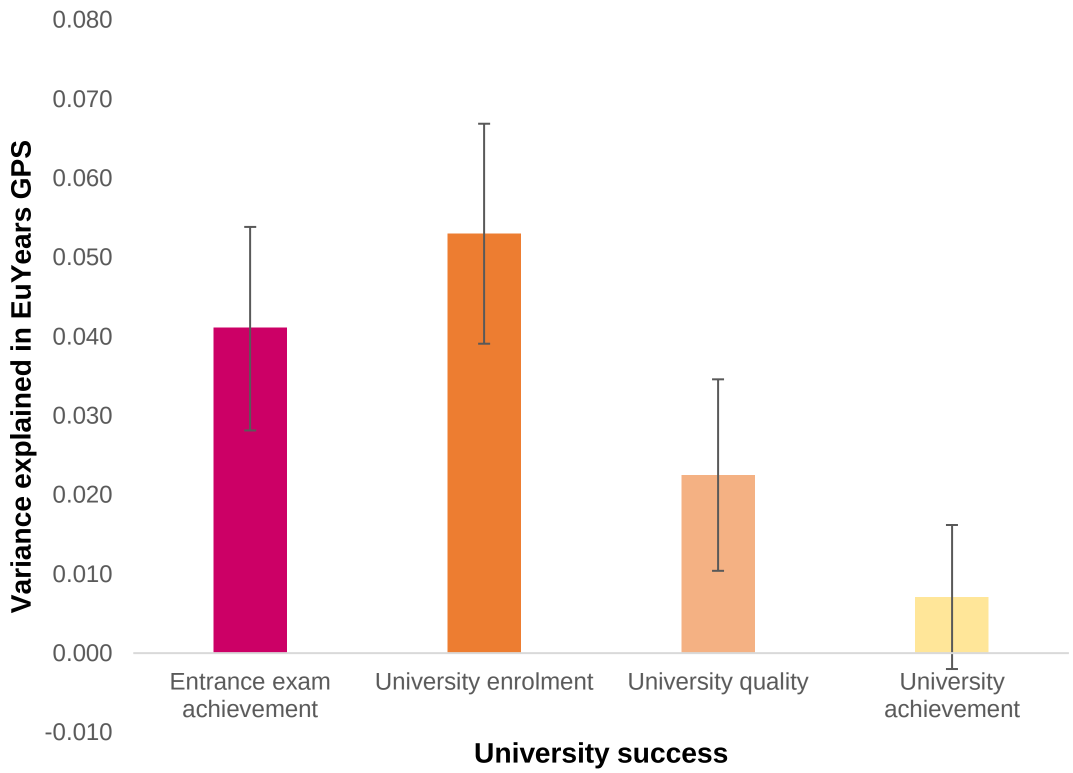


Figure 2 – Variance explained (R^2) and 95% confidence intervals by *EduYears* genome-wide polygenic score for each of the university success variables.

Discussion

Our results represent the first genetically sensitive exploration of success at university using twin and genomic data. Twin analysis revealed substantial heritability for all university success measures, including university entrance exam achievement (57%), the choice to study at university (51%), the quality of university attended (57%) and achievement at university (46%). In addition to twin analysis, we also found evidence for genetic influence using DNA alone. Indeed, a genome-wide polygenic score (GPS) for adult educational attainment (Okbay et al., 2016) explained up to 5% of variance in the university success variables. Taken together, these results highlight that the appetite and aptitude young adults have for higher education is, in part, genetically influenced.

Finding genetic influence on success in university extends a vast literature on education and genetics (Asbury & Plomin, 2013; Kovas, Haworth, Dale, & Plomin, 2007). The present results show for the first time that genetic influence on educational achievement continues to university. This is in line with twin estimates in earlier school years. For example, one study (Rimfeld et al., 2016) using the same sample found that at age 18, the heritability of achievement in different subjects ranged from 23-82%. Interestingly, the substantial influence of the shared environment on educational achievement during the early school years tapers off at university. Indeed, shared environmental influences account for up to 20% of the variance in the compulsory school years (Knopik, Neiderheiser, DeFries, & Plomin, 2017), but make a non-significant contribution to variance in achievement at university. One explanation for this pattern of results is that in the early school years children's environments are largely the same across multiple life domains, for example siblings go to the same school, have many of the same friends, and spend much of their time at home under substantial parental influence. By contrast, young adults have more freedom of choice in their education, in terms of the subjects they take, the extracurricular activities they engage in and how they spend their time. This increase in choice leads to greater genetic influence and decreased shared environmental influence across development. We see this same developmental decrease in the influence of the shared environment for other educationally-relevant traits such as intelligence (Plomin & Deary, 2015). This suggests that as children gain more freedom to choose their environments, they increasingly select environments that correlate with their genotype.

An exception to this developmental decrease in shared environmental influence pertains to decisions about whether to continue in education. For example, shared environmental influences account for nearly 40% of the variance in the choice of whether or not to take A-levels (Rimfeld et al., 2016) and the choice of whether or not to pursue a university degree, yet shared environmental influence is not evident for university achievement. It is possible that families and schools influence educational choices to a greater extent than educational achievement.

Multivariate genetic analyses indicated a substantial genetic correlation between university entrance exam achievement and university quality (76%). These results support the generalist genes hypothesis of cognitive traits (Davis, Haworth, & Plomin, 2009), which suggests that most genetic influences are shared across learning abilities and therefore educationally relevant genes will influence a range of associated traits, for example intelligence (Selzam et al., 2017), SES (Selzam et al., 2017) and now, university success. Along with genetic correlations, we also found moderate shared environmental correlations between university success variables (r_c = 27-81%). Non-shared environmental influences were mainly uncorrelated (r_e= 3-35%). This suggests that unique environmental factors that contribute to variance in university success are idiosyncratic and time specific and do not contribute to effects across compulsory and higher education.

Although we found moderate twin heritability estimates for university achievement, the polygenic score prediction of this trait was small in magnitude, only predicting 0.7% of the variance. Furthermore, even when we split university achievement into two groups: STEM-related subjects and humanities subjects, we did not find any differences in *EduYears* GPS prediction between subjects. This suggests that even within subject field, the GPS is not discriminative of achievement. In contrast, *EduYears* GPS explains 9% of the variance in achievement at age 16 (Selzam et al., 2017). There are several possible reasons for these ostensibly conflicting results. First, a polygenic score based on years of education might be less discriminative for individuals who have all obtained a university degree. Second, examinations at university level are not standardised, which means that results may be less comparable between universities; a first-class degree at an elite university will be weighted the same as one from a lower-level university. This interpretation is supported by the low MZ correlations for university achievement (0.30), compared to the MZ correlations for the other university success measures (0.50-0.69). Such coarseness in measurement may render the *EduYears* polygenic score less capable of predicting individual differences. Finally, it is possible that getting into university and achievement at university are predicted by different heritable traits. Indeed even standardised tests such as the Scholastic Aptitude Tests (SATs) that are widely used for college admittance in the United States are poor predictors of both four and six-year university graduation rates after admittance (Bowen, Chingos, & McPherson, 2009). Future studies using multivariate genetic modeling can test this differential heritability hypothesis.

In contrast to the results for university achievement, *EduYears* polygenic score predicted variance in both the decision to attend university, as well as the choice of which university to attend. These results are in line with our twin analysis demonstrating substantial genetic influence on educational choices. Both the decision of whether or not to go to university, and which university to attend, are influenced by an individual's educational qualifications, which we know are substantially heritable (Shakeshaft et al., 2013). However, even once we controlled for prior academic achievement, the quality of university attended was still considerably heritable (47%). This is likely because, in addition to getting the right grades, there are other heritable factors which influence both the decision to go to university, as well as the decision to go to one university over another, for example socio-economic status, friendships, secondary school quality and parental involvement with students' learning (Frenette, 2007).

The present study benefited from a large sample size of over 3,000 twin pairs and over 3,000 genotyped individuals, as well as a multi-method approach. However, our results must be considered in light of limitations of current DNA methods, in addition to the general limitations of the twin method (Knopik et al., 2017).

The *EduYears* GPS explains only a fraction of the known high heritability of educationally relevant traits as estimated from twin studies. This is because GPS are derived from GWA studies that are limited to estimating additive genetic effects from common SNPs present on DNA arrays or variants in linkage disequilibrium. For this reason, GPS will underestimate genetic influence to the extent that non-additive effects or rare variants contribute to its heritability. However, there has been limited success for detecting non-additive variation in GWA studies. Potential reasons for this are 1) non-additive effects do not appear to make up a large fraction of the total genetic variation, as identified by twin studies and 2) because effects are likely very small, large sample sizes would be needed (Visscher et al., 2017). SNP-based estimates of heritability, which have these same limitations, represent the current upper limit for GPS prediction. Although we were underpowered to calculate SNP-based heritability estimates in the present study, our data collection is ongoing, and we plan to explore SNP-based estimates for university success in the future. As the so-called missing-heritability gap closes, GPS predictions will improve and will increasingly be used as an index of genetic influence on complex human behavior (Plomin, in press).

Despite this limitation of our molecular genetic analysis, this study represents the first genetically informative study of university success. We show that genetic influences on education trajectories are pervasive and cumulative into young adulthood and affect both appetite for education and aptitude for learning.

Method

Participants

Participants were drawn from the UK-representative Twins Early Development Study (TEDS). TEDS is a multivariate and longitudinal birth cohort study that recruited over 15,000 twin pairs born in England and Wales between January 1994 and December 1996. The representativeness of the TEDS sample has been assessed longitudinally and is described in further detail elsewhere (Haworth, Davis, & Plomin, 2013; Kovas, Haworth, Dale, & Robert, 2007). The TEDS twins are representative of the UK population for each of our university success variables. For percentages of TEDS participants were similar to UK national averages for enrolling in university (56% vs 49%) and obtaining a first class degree (33% vs 26%) (Department for Education, 2017b; Higher Education Statistics Agency, 2017). In addition, the genotyped sub-sample is representative of the UK for gender, parental education and rates of employment for both mothers and fathers (Selzam et al., 2017).

All analyses were conducted on participants without severe neonatal problems. Ethical approval for this study was received from King's College London Ethics Committee and all methods were carried out in accordance with the relevant guidelines and regulations. All participants provided written informed consent.

Twin sample

Zygosity was based on parent reports of twin differences during childhood, which is over 95% accurate when compared to DNA testing (Price et al., 2000). For cases where zygosity was unclear, DNA testing was conducted. After exclusions, data on entrance exam achievement were available for 4,698 twin pairs (9,407 individuals), of which 3,339 were monozygotic (MZ) twin pairs, 3,040 were dizygotic (DZ) same-sex twin pairs and 3,028 were DZ opposite-sex twin pairs. Data on university enrolment were available for 5,143 twin pairs (10,288 individuals), of which 3,591 were MZ twin pairs, 3,364 were DZ same-sex twin pairs and 3,333 were DZ opposite-sex twin pairs.

Data on university quality were available for 2,948 twin pairs (5,941 individuals), of which 2,086 were MZ twin pairs, 1,882 were DZ same-sex twin pairs and 1,973 were DZ opposite-sex twin pairs. Finally, data on university achievement were available for 1,590 twin pairs (3,219 individuals), of which 1,222 were MZ twin pairs, 985 were DZ same-sex twin pairs and 1,012 were DZ opposite-sex twin pairs

Genomic sample

The TEDS sample includes a genotyped subsample of unrelated individuals (i.e., one member of a twin pair). Genotypic analyses were restricted to participants of European decent, as ascertained by TEDS questionnaire data at first contact when the twins were aged 2. Standard principal component analyses were used to confirm the European ancestry of the sample. Here we regressed the GPS on the first 10 principal components and used the residuals in all subsequent analyses. This procedure controls for population stratification, which is the systematic difference in allele frequencies observed in subpopulations of individuals of different ancestry. Genomic data for creating genome-wide polygenic scores (GPS) were available for 3,501 individuals with data on entrance exam achievement, 3,774 individuals with data on university enrolment, 2,251 individuals with data on university quality and 1,291 individuals with data on university achievement. This genotyped sample is representative of UK census data on education and socioeconomic related phenotypes for families with children, for example the percentage whose parents went on to further education and parental employment (Selzam et al., 2017).

DNA was genotyped using Illumina HumanOmni ExpressExome-8v1.1 arrays (Institute of Psychiatry, Psychology and Neuroscience Genomics & Biomarker Core Facility, London, United Kingdom) or Affymetrix GeneChip 6.0 DNA arrays (Affymetrix, Santa Clara, CA). The sample with genotype data consisted of 5,825 individuals (2,698 genotyped with Illumina and 3,127 genotyped with Affymetrix arrays). Genome wide genotypes from the two arrays were separately imputed using the Haplotype Reference Consortium (Haplotype Reference Consortium, 2016) and the imputation software Minimac3 1.0.13 (Fuchsberger, Abecasis, & Hinds, 2014), which are available from the Michigan Imputation Server (https://imputationserver .sph.umich.edu). A series of quality checks were performed before merging data from the two arrays imputation (e.g. array effects, allele frequencies by imputation quality). For the present analyses, we limited our analyses to variants genotyped or imputed at info >0.95 on both arrays, and with Hardy Weinberg Equilibrium test *p*-value $>10^{-5}$.

Linkage disequilibrium (LD) refers to the non-random association of alleles at different loci. When calculating polygenic scores using PRSice those markers in high LD are removed or 'pruned' as correlated variants can represent non-independent association signals that if ignored can overweight GPS in favour of loci in high LD. Stringent pruning resulted in the exclusion of eight genomic regions in high linkage disequilibrium ($R^2 > 0.1$ cutoff within a 250-kb window). To ensure that only genome wide effects were detected, we performed a principal components analysis to correct for possible stratification using a subset of 40,745 autosomal single-nucleotide polymorphisms (SNPs) that remained after we applied our quality-control criteria and that overlapped between the two genotyping arrays.

Measures

The present study included individuals with data available on measures of 'university success', as described below. These measures were obtained from the twins at ages 18 and 22 using paper and online questionnaires, as well as a mobile phone application. For sample sizes across each of the measures, please see Table S1.

University entrance exams variables

In the UK, getting into university is dependent on achievement on the General Certificate of Education Advanced Level, or 'A-levels' (https://www.ucas.com/ucas/undergraduate/getting-started/ucas-undergraduate-entry-requirements). A-levels are a two-year school leaving qualification offered at the end of compulsory education at age 16, when students are allowed to decide, for the first time, whether they want to continue formal education. In addition to choosing whether they wish to continue education, students are also able to pick what they want to study, with students typically completing three A-levels in a variety of subjects. If students only complete one out of the two A-level years they are awarded an AS-level qualification.

Entrance exam achievement

A-levels are graded from E, the minimum pass grade to A*, the best possible grade. For university entry, these grades are converted into Universities and Colleges Admissions Service (UCAS) points. Following this point system, achieving an A* at A-level is awarded 56 points, whereas an E is awarded 16 points. If students only complete an AS-level qualification, UCAS points are adjusted accordingly, with an A* grade at AS-level awarded 20 points and an E grade at AS-level awarded 6 points (for further information on UCAS points, see: https://www.ucas.com/ucas/tariff-calculator).

Data on UCAS point scores were available for our sample through the National Pupil Database: [https://www.gov.uk/government/collections/national-pupil-database.](https://www.gov.uk/government/collections/national-pupil-database)

University variables

University enrolment

Data on whether or not twins chose to attend university was collected via questionnaire at age 18. This questionnaire was designed to assess post-18 destinations. Choosing to attend university was treated as a dichotomous variable, where 1 indicated the choice to pursue university and 0 indicated any other post compulsory education destination, such as going into employment, training or unemployment. Approximately 57% of our sample reported accepting a place at university, which is similar to the UK average (Department for Education, 2017a).

University quality

For those who indicated that they were attending university, we also asked for the name of the university they attended. We used this information to create a university quality measure by ranking the universities in order based on the UK university league tables in 2014 (the year that the majority of the sample applied to university) (The Complete University Guide). This ranking system takes into account the entry standards of the university, the average UCAS points of students at the university, research output, and graduate prospects. According to this ranking system the University of Cambridge was at the top, and East London University was at the bottom, with 124 universities in total.

To explore the unique contribution of our university quality variable beyond the effect of previous educational achievement, we regressed university quality on entrance exam achievement.

University achievement

At age 22 (M = 22, SD = 0.85), we contacted twins about their higher education choices, including whether or not they had completed an undergraduate degree and what grade they obtained. Undergraduate university degrees were graded from 1 (the lowest possible pass) to 5 (a first-class degree, the highest possible pass). Of those individuals who indicated that they were attending university at age 18 (N = 5,833), over half provided their final university grade at age 22 (N = 3,219). To check whether there were any achievement or university quality differences between those who reported their final degree grade and those who did not, we performed sensitivity analysis. Here, a t-test compared entrance exam achievement and university quality between those individuals who stated they were attending university but did or did not report their final university degree grade (see Supplementary Table S2).

Degree subject

At this age, we also asked twins to choose a category that best described their degree subject (see Table S13 for a list of categories and sample sizes). We classed those people who selected ‘natural sciences’, ‘mathematics/statistics’, ‘medicine/veterinary’, ‘engineering’, ‘technology/design’ or ‘computing/IT’ as STEM and those who took ‘social sciences’, ‘arts’, ‘humanities’, ‘languages’ or ‘law’ as Humanities. Descriptive statistics for STEM vs Humanities degree subjects can be found in Table S13.

A correlation matrix of all of the variables included in the current study can be found in Supplementary Figure S1

Statistical analyses

Twin analyses

Univariate twin analyses were used to compare twin similarity on university entrance examinations, university quality and university achievement between MZ twin pairs who share 100% of their genetic material and DZ twin pairs who share on average 50% of the genetic material that can differ between individuals (segregating alleles). Twin analyses were used to estimate the proportion of variance in university entrance examinations, university quality and university achievement that can be attributed to genetic and environmental factors (Knopik et al., 2017). The genetic contribution to phenotypic variance is referred to as heritability (A) and is narrowly defined as the proportion of individual differences in a population that can be attributed to additive effects of inherited DNA differences between individuals. We can roughly estimate ‘A’ by doubling the difference between MZ and DZ intraclass correlations on a trait. The environmental contribution to phenotypic variance includes those non-inherited influences that are shared (C) and unique (E) to twins growing up in the same home. The C component refers to those environmental factors that contribute to twin similarity and can be calculated by subtracting ‘A’ from the MZ twin correlations. The E component captures environmental experiences unique to the individual as well as measurement error and is measured by deducting the A and C components from unity.

The ACE estimates for twin analyses can be calculated more precisely using structural equation modelling (SEM) with the OpenMX software package (Boker et al., 2011), which also provides confidence intervals around the estimates.

Statistical approaches for analyzing twin data are described elsewhere (Neale & Cardon, 1994; Rijsdijk & Sham, 2002). Briefly, SEM leverages the different sources of sibling similarity and differences to make inferences on the aetiology of observable traits. SEM tests hypotheses about relations among observed phenotypic correlations and latent genetic and environmental factors by modeling the observed covariance between MZ and DZ twin pairs on the phenotype. Here, model parameters are estimated by minimising a goodness-of-fit statistic that seeks to obtain the smallest possible discrepancy between the model and the observed data. A likelihood ratio chi-square statistic (χ^2) is then used to measure the goodness of fit of the tested model relative to a perfectly fitting (saturated) model. A significant (p <.05) χ^2 when comparing the tested model to the saturated model means that the model provides a poor fit to the data and can be rejected, while a non-significant χ^2 means that the model is consistent with the data. In the present analyses, we tested a series of nested models to determine whether the components, A, C and E, are significantly greater than zero. With each test we assessed whether the fit of the simpler, nested model was significantly worse than that of the full model, with preference for a simpler more parsimonious explanation of the observed data. Details of each model tested are presented in Appendix 4.

Finally, a liability threshold model (LTM) was used to compute ACE estimates for university enrolment. The LTM is an extension of the classic univariate twin analyses, used for dichotomous variables, for example, in case-control studies comparing individuals with diagnoses to those without. Here, binary variables are assumed to represent an unobserved normal distribution (Boker et al., 2011) and twin tetrachoric (rather than intraclass) correlations are compared to index relative genetic and environmental contribution to the liability. Similar to the univariate model, greater MZ compared to DZ correlations can be used to estimate the ACE components to the liability variance. Sub models comparisons for the LTM SEM compared a fully saturated model with a constrained model (Sub 1) where thresholds were constrained across twin 1 and twin 2 within zygosity groups and a second model (Sub 2) where thresholds were equated across twin pairs and zygosity. Similar to assessment of univariate and multivariate SEM described above, model fit statistics were used to isolate the most parsimonious fit to the data.

Multivariate model fitting is an extension of univariate twin analyses that relies on cross-twin cross-trait correlations to decompose phenotypic covariance between multiple traits into genetic and environmental components of covariance. The correlated factor model, which was found to be the best fit to the data, was used to estimate A, C, and E correlations between our continuous university success variables. This model assumes each variable is influenced by a set of genetic, shared and non-shared environmental factors that are allowed to correlate with each other through r_A, r_C and r_E. Model fit statistics are provided in the Supplementary Online Material (Table S10).

Genomic analyses

A genome-wide polygenic score (GPS) was derived from summary statistics from a published genome-wide association (GWA) study of years of education (Okbay et al., 2016). The GPS serves as an individual-specific genetic prediction derived directly from DNA and is calculated by summing genotypic values for each trait-associated single nucleotide polymorphism (SNP) weighted by its association in the GWA study sample. A GPS was calculated for each of the unrelated, genotyped individuals in the TEDS sample using PRSice (Euesden, Lewis, & O'Reilly, 2014). Here, PRSice performs a regression analysis to test for association between GPS and each of our university success outcomes. We used the high-resolution scoring option in PRSice which calculates GPS at a large number of *p*-value thresholds, ranging from 0.001 to 1 (increments of 0.001) in the GWA study results. For all university success variables, the most predictive threshold was 0.05 (i.e., including all GWA study identified SNPs with *p*-value up to 0.05), which included 19,415 SNPs. The results for the 0.05 threshold are reported in the manuscript while results for each of the other tested thresholds is provided in Figure S2.

The difference between genetic estimates from twin and polygenic score analysis should be noted. Additive genetic influences (A) in twin models include additive genetic effects of any DNA sequence differences. In contrast, the polygenic score prediction only includes the additive effects of common SNPs in DNA arrays that have been linked to the target trait of a GWA study. Therefore, it is to be expected that GPS will account for a small amount of variance compared to the sum total of all genetic effects (heritability) estimated in twin studies.

Regression models were then used to estimate the proportion of variance in the continuous (R^2) or dichotomous (Nagelkerke R^2) university success variables that can be explained by variance in individuals' GPS. Furthermore, to test for potential correlation differences between *EduYears* GPS predictions of university degree grades for STEM versus humanities subjects, we performed Fisher's *r*-to-*z* transformations.

Data availability

For information on data availability, please see the Twins Early Development Study data access policy. This can be found at: http://www.teds.ac.uk/research/collaborators-and-data/teds-data-access-policy.

Author Contributions

ESW, ZA and RP conceived and designed the study. ESW and ZA analysed the data and wrote the manuscript. All the authors discussed the results and implications and commented on the manuscript at all stages.

Competing Interests

The authors declare that they have no conflicts of interest with respect to their authorship or the publication of this article.

Acknowledgements

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a programme grant to R.P. from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ grant agreement n° 602768 and ERC grant agreement n° 295366. RP is supported by a Medical Research Council Professorship award (G19/2). SvS is supported by a Jacobs Foundation Research Fellowship Award (2017–2019).

References

Asbury, K., & Plomin, R. (2013). *G is for genes: The impact of genetics on education and achievement* (Vol. 24): John Wiley & Sons.

Ayorech, Z., Plomin, R., & von Stumm, S. (in press). Using DNA to predict educational choices in early adulthood. *Developmental Psychology*.

Ayorech, Z., von Stumm, S., Haworth, C. M., Davis, O. S., & Plomin, R. (2017). Personalized media: a genetically informative investigation of individual differences in online media use. *PloS one, 12*(1), e0168895.

Baker, L. A., Treloar, S. A., Reynolds, C. A., Heath, A. C., & Martin, N. G. (1996). Genetics of educational attainment in Australian twins: sex differences and secular changes. *Behavior genetics, 26*(2), 89-102.

Bartels, M., Rietveld, M. J., Van Baal, G. C. M., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research and Human Genetics, 5*(6), 544-553.

Belsky, D., Domingue, B., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., . . . Mullan, H. K. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1801238115

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., . . . Bates, T. (2011). OpenMx: an open source extended structural equation modeling framework. *Psychometrika, 76*(2), 306-317.

Bolinskey, P. K., Neale, M. C., Jacobson, K. C., Prescott, C. A., & Kendler, K. S. (2004). Sources of individual differences in stressful life event exposure in male and female twins. *Twin Research and Human Genetics, 7*(1), 33-38.

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at America's public universities*: Princeton University Press.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual review of psychology, 53*(1), 371-399.

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social forces, 92*(1), 109-140.

Brooks, R. (2005). Measuring university quality. *The Review of Higher Education, 29*(1), 1-21.

Carnevale, A. P., Cheah, B., & Hanson, A. R. (2015). The economic value of college majors. *Georgetown University Center on Education and the Workforce*.

Davis, O. S., Haworth, C. M., & Plomin, R. (2009). Learning abilities and disabilities: Generalist genes in early adolescence. *Cognitive neuropsychiatry, 14*(4-5), 312-331.

Department for Education. (2017a). Participation Rates In Higher Education: Academic Years 2006/2007 – 2015/2016. from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/648165/HEIPR_PUBLICATION_2015-16.pdf

Department for Education. (2017b). Participation Rates In Higher Education: Academic Years 2006/2007 – 2015/2016. Department for Education, UK.

Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher education, 49*(4), 495-533.

Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2014). PRSice: polygenic risk score software. *Bioinformatics, 31*(9), 1466-1468.

Falconer, D. S. (1960). *Introduction to Quantitative Genetics*. New York: Ronald Press.

Frenette, M. (2007). *Why Are Youth from Lower-Income Families Less Likely to Attend University?: Evidence from Academic Abilities, Parental Influences, and Financial Constraints*. Ottawa, Canada: Statistics Canada Ottawa.

Fuchsberger, C., Abecasis, G. R., & Hinds, D. A. (2014). minimac2: faster genotype imputation. *Bioinformatics, 31*(5), 782-784.

Fulker, D. W., & Eysenck, H. (1979). Nature and nurture: Heredity *The structure and measurement of intelligence* (pp. 102-132): Springer.

Haplotype Reference Consortium. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics, 48*(10), 1279-1283.

Haworth, C. M., Davis, O. S., & Plomin, R. (2013). Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics, 16*(1), 117-125.

Higher Education Statistics Agency. (2017). Higher Education Student Statistics: UK, 2016/17 - Qualifications achieved. London, UK.

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics, 91*(4), 717-724.

Jaffee, S. R., & Price, T. S. (2007). Gene–environment correlations: A review of the evidence and implications for prevention of mental illness. *12*(5), 432-442.

Jones, E. B., & Jackson, J. D. (1990). College grades and labor market rewards. *The Journal of Human Resources, 25*(2), 253-266.

Knopik, V. S., Neiderheiser, J., DeFries, J. C., & Plomin, R. (2017). *Behavioral genetics* (7th Ed.). New York: Worth.

Kovas, Y., Haworth, C., Dale, P. S., & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for research in Child Development, 72*(3), vii, 1-144.

Kovas, Y., Haworth, C., Dale, P. S., & Robert, P. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for research in Child Development, 72*(3), vii, 1-144.

Neale, M. C., & Cardon, L. R. (1994). Methodology for genetic studies of twins and families. *STATISTICS IN MEDICINE, 13*, 199-199.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Meddens, S. F. W. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539-542.

Plomin, R. (1994). *Genetics and experience: The interplay between nature and nurture*: Sage Publications, Inc.

Plomin, R. (in press). *Blueprint: How DNA makes us who we are*. London: Allen Lane (Penguin).

Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influence on “environmental” measures. *Behavioral and Brain Sciences, 14*(3), 373-386.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular psychiatry, 20*(1), 98.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin, 84*(2), 309.

Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research and Human Genetics, 3*(3), 129-133.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in higher education, 16*(2), 129-150.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., . . . Agrawal, A. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *science, 340*(6139), 1467-1471.

Rijsdijk, F. V., & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in bioinformatics, 3*(2), 119-133.

Rimfeld, K., Ayorech, Z., Dale, P. S., Kovas, Y., & Plomin, R. (2016). Genetics affects choice of academic subjects as well as achievement. *Scientific reports, 6*, 26373.

Rimfeld, K., Malanchini, M., Krapohl, E., Hannigan, L. J., Dale, P. S., & Plomin, R. (in press). The stability of educational achievement across school years is largely explained by genetic factors. *npj Science of Learning*.

Saudino, K. J., Pedersen, N. L., Lichtenstein, P., McClearn, G. E., & Plomin, R. (1997). Can personality explain genetic influences on life events? *Journal of personality and social psychology, 72*(1), 196-206.

Selzam, S., Krapohl, E., von Stumm, S., O'reilly, P., Rimfeld, K., Kovas, Y., . . . Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular Psychiatry, 22*(2), 267-272.

Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C. M., . . . Plomin, R. (2013). Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PloS one, 8*(12), e80341.

Smith-Woolley, E., Pingault, J.-B., Selzam, S., Rimfeld, K., Krapohl, E., von Stumm, S., . . . Plomin, R. (2018). Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them. *npj Science of Learning, 3*(1), 3.

Tam, M. (2001). Measuring quality and performance in higher education. *Quality in higher Education, 7*(1), 47-54.

The Complete University Guide. University League Tables 2014. from https://www.thecompleteuniversityguide.co.uk/league-tables/rankings?y=2014

Van Der Waerden, B. L. (1975). On the sources of my book Moderne Algebra. *Historia mathematica, 2*(1), 31-40.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics, 101*(1), 5-22.

Chapter 6 – General discussion, implications and future directions

The aim of this thesis was to better understand the genetic and environmental contributions to academic achievement at secondary school level and beyond. This thesis focused on both the characteristics that students bring to the school (their unique set of genes, personality, ability, prior achievement and socioeconomic status), as well as school-wide environments (school type and school quality) and investigated how they influence individual differences in academic achievement. This chapter summarises the key findings from the preceding chapters and discusses limitations, implications and future directions.

SUMMARY OF FINDINGS

Previous research using twin studies has shown that academic achievement throughout education is highly heritable and influenced by many factors including general cognitive ability, personality, and the school environment. This thesis investigated these factors and their relationship to academic achievement at the end of compulsory education and into university.

In Chapter 2, the relationship between personality and achievement was explored using a genome-wide polygenic score (GPS) derived from a genome-wide association (GWA) study of years of education (*EduYears*). It was shown that DNA variants contributing to how long individuals stay in education are also predictors of personality, explaining up to 2% of the variance. This finding remained significant even after controlling for general cognitive ability. Furthermore, it was found that *EduYears* GPS was a better predictor of personality domains than personality GPS themselves. Using structural equation modelling, it was shown that *EduYears* GPS explained between 7-16% of the covariance between personality and educational achievement at the end of compulsory education at age 16. These results demonstrate the substantial genetic pleiotropy across educational achievement and educationally relevant traits.

Chapter 3 investigated a hotly debated school-wide environment: school type. This thesis found that the considerable achievement differences between students attending selective schools (private and grammar school) compared to those attending non-selective schools, were mainly explained by the factors schools use in selection.

Furthermore, these same factors explained most of the average *EduYears* GPS difference between students attending different schools. These results demonstrate that genetic and exam differences between school types are primarily due to the heritable characteristics involved in pupil admission.

Another school-wide factor thought to explain differences in academic achievement at the end of compulsory education is Ofsted-rated school quality. Parents use this publicly available school quality rating as a means to help them decide on a secondary school for their children. Chapter 4 found that Ofsted-rated school quality explains 4% of the variance in academic achievement at the end of compulsory education. However this reduces to less than 1% after accounting for student characteristics (family socioeconomic status and prior achievement). Furthermore, Ofsted-rated school quality was not a significant predictor of wellbeing or school engagement. In other words, students attending 'Outstanding' schools reported similar levels of wellbeing, bullying and aspirations compared to those attending 'Inadequate' schools, on average. These findings call into question the usefulness of Ofsted ratings for parents and students when choosing secondary schools.

Much of the previous behavioural genetic literature estimating the aetiology of academic achievement has focused on secondary school education, with no twin studies looking at the aetiology of academic achievement at university level. Chapter 5 uses a multi-method approach to investigate the genetic basis of university success, comprising: achievement on university entrance examinations, university enrolment, university quality, and university achievement. This chapter found that inherited DNA differences explain up to 57% of the individual differences in university success variables. Furthermore, multivariate twin analysis revealed that a large proportion of the correlations between university success measures are explained by genetic factors. The twin analysis is complemented by analysis using *EduYears* GPS which explained up to 5% of the variance in the university success variables. These findings provide support for continued and substantial heritability of academic achievement into higher education.

LIMITATIONS

The limitations of the twin design and of GPS are discussed in the general Methods section of Chapter 1, and the specific limitations of each of the studies presented in this thesis are considered in earlier chapters. This section discusses three general limitations relevant to the present thesis: 1) the use of self-report data; 2) the possible effect of primary schooling; and 3) generalisability to different schooling systems.

Use of self-report data

Much of the data used in the present thesis was self-reported by the twins. For example: personality and motivation (Chapter 2), school type (Chapter 3), wellbeing and school engagement (Chapter 4), university enrolment (Chapter 4) and in many cases, academic achievement (all chapters). Although self-report information can be useful to obtain an individual's perspective on their own behaviour, it can also be unreliable due to social desirability effects or faking. Faking describes deliberately giving false information to create a different impression or persona (Furnham, 1986). A more specific facet of this is social desirability bias, in which an individual distorts self-reports favourably to come across more positively (Nederhof, 1985; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Social desirability and faking responses have been found for personality questionnaires, psychiatric symptom inventories, and questionnaires asking about motivation (Furnham, 1986). A way around this could be to use multiple raters, such as parents, siblings or teachers. However, correlations between self and parent-report are often low, especially for questionnaires that require the judgement of internal processes (Verhulst & Van der Ende, 1992). Furthermore, the structure of genetic and environmental influence on personality and psychiatric traits tend to differ between raters (Kendler, Prescott, Myers, & Neale, 2003). However, with regards to personality, wellbeing and school engagement, who better to report on this information than twins themselves? To best understand how these personality traits link to academic achievement, getting self-report data is the most useful.

In the present thesis the reliability of self-reported academic achievement at age 16 (General Certificate of Secondary Education: GCSEs) and school type data was checked by comparing data collected by the Twins Early Development Study (TEDS) to data collected by the National Pupil Database (NPD). GCSE results collected from twins correlated at .99 with those collected by the NPD. For school type, over 75% of individuals correctly gave their school type (see Chapter 2 for further details). This comparison indicates that, at least for achievement and school type measures, self-reporting was not heavily influenced by bias.

Effects of primary schooling

Another limitation of the present thesis is that it only focused on outcomes in adolescence and early adulthood. However, school environments could influence academic outcomes differently in the early years. For example, a meta-analysis of student-teacher relationships using 61 studies and a combined sample size of over 88,000 found that negative teacher-student relationships appeared to have a slightly larger negative impact on students' engagement and achievement in primary school (-.34 and -.19 respectively) compared to secondary school (-.25 and -.13 respectively) (Roorda, Koomen, Spilt, & Oort, 2011). In terms of school type, one study found that private primary schooled children had higher reading and mathematics scores at ages four and eight compared to their state primary school educated peers, even after accounting for family socioeconomic status, gender and the average achievement of the children in the school (Ndaji, Little, & Coe, 2016). However, although these studies suggest that school environments may have a larger impact on children's achievement in the early years, evidence from behavioural genetic studies suggests that environments are unlikely to explain the stability in academic achievement across development.

A recent study (Rimfeld et al, in press) looking at the genetic and environmental influences on academic achievement from primary school (age seven) to the end of secondary school (age 16), found that achievement was highly stable (phenotypic correlations ~.70) but that much of this stability was accounted for by genetics. The environment, on the other hand, specifically the non-shared environment, explained most of the age-to-age change in achievement scores. This is in line with the general rubric of behavioural genetics: 'genetics accounts for the stability, environments account for the change' (Deary et al., 2012; Knopik, Neiderheiser, DeFries, & Plomin, 2017; Plomin, DeFries, Knopik, & Neiderhiser, 2016). If this is indeed the case for academic achievement, then environments such as school type and school quality are unlikely to explain changes in achievement across time as these environments are typically shared between twins.

Generalisability to other education systems

A further limitation of the present thesis concerns generalisability. The research presented in this thesis uses data from the English education system at a specific point in time and may therefore not be generalisable to other education systems around the world. Even within the English education system, changes are made to the curriculum and to examination scoring on a regular basis. For example, participants in TEDS completed their General Certificate of Secondary Education (GCSEs) at age 16, GCSE scoring has partially changed. Now, instead of being graded from G (the lowest possible grade) to A* (the highest possible grade), some GCSEs are scored from 1 (low) to 9 (high). This reform will be phased over a number of years and will take until 2020 to complete. This reform aims to reduce the GCSE ceiling effect by discriminating more at the top end. Figure 2 shows the proportions of students at each grade in GCSE English and maths in 2016, under the old A*-G scoring system, compared to 2017 under the new 1-9 scoring. It is possible to see that, at the top end under the new system, grade 9 captures the top 2-3% of students, whereas grade A*, under the old system, captured 4-7%. Once the new scoring system has been phased in for all GCSE subjects, it will be useful to re-examine the effects of school-wide systems such as school type and Ofsted, which may potentially explain more variance.

However, even with big changes to the education system, it is interesting to note that the variance in achievement explained by school type has not changed much. For example, a review of school type effects on educational achievement conducted with data from the National Child Development Survey (Sullivan & Heath, 2002), found very similar results to the results reported in Chapter 3. This is despite the fact that this data was collected in the 1950-1970s, during the Tripartite System (children were sent to grammar, secondary modern or technical schools based on the outcome of a test at age 11), rather than the comprehensive education system we have today.

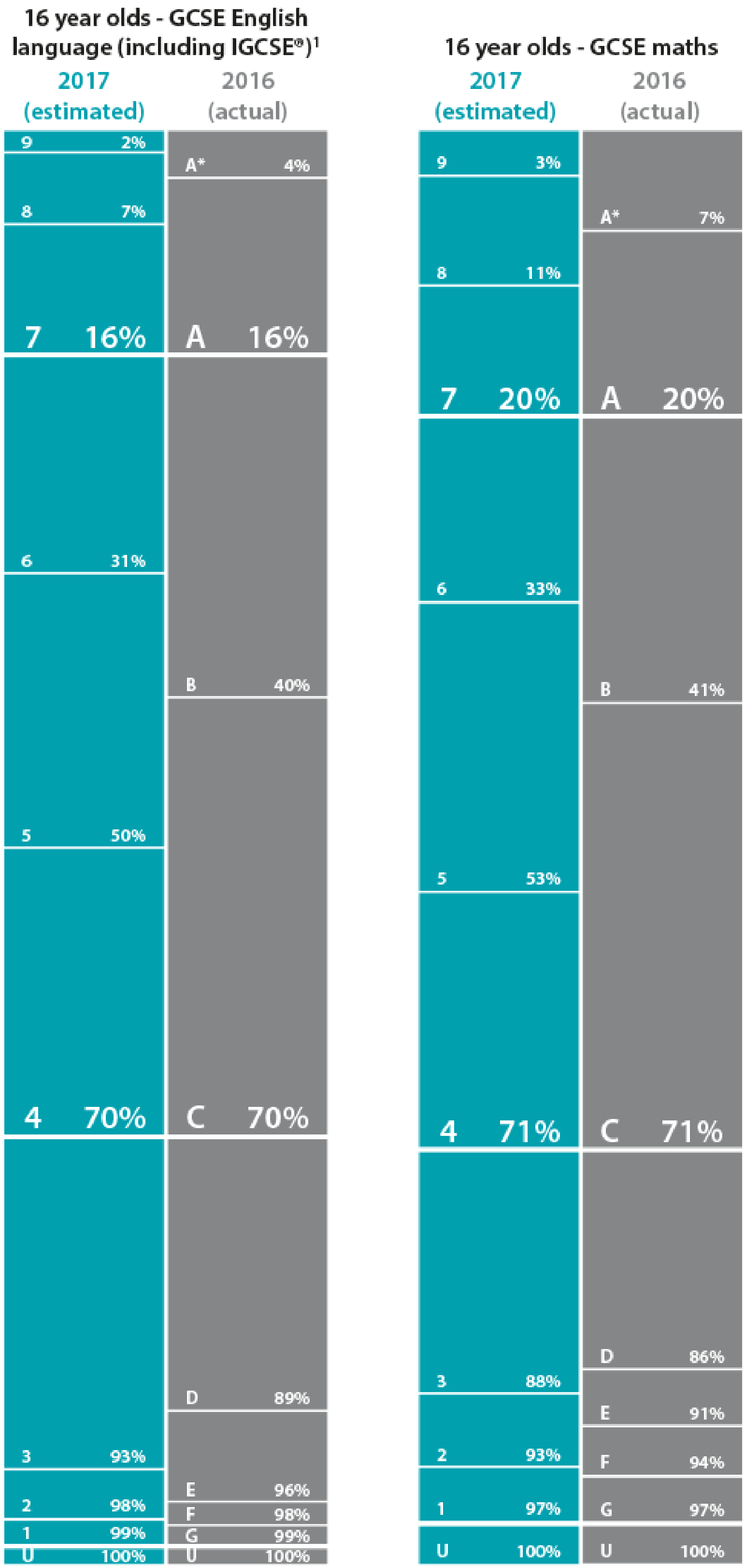


Figure 2 – Estimates of proportions of students at each grade in GCSE English language and maths in 2017 (16 year olds, England only), Ofqual (https://ofqual.blog.gov.uk/2017/04/05/setting-grade-9-in-new-gcses/)

In terms of the generalisability to other countries, although education systems differ substantially across the world, there is a great deal of overlap in achievement between countries (Organisation for Economic Co-operation and Development, 2017). Figure 3 shows the percentage of students obtaining mathematics level 1 (can answer basic questions where all relevant information is present and the questions are clearly defined) to level 6 (students are capable of advanced mathematical thinking and reasoning) across 70 countries. These grades overlap substantially between countries.

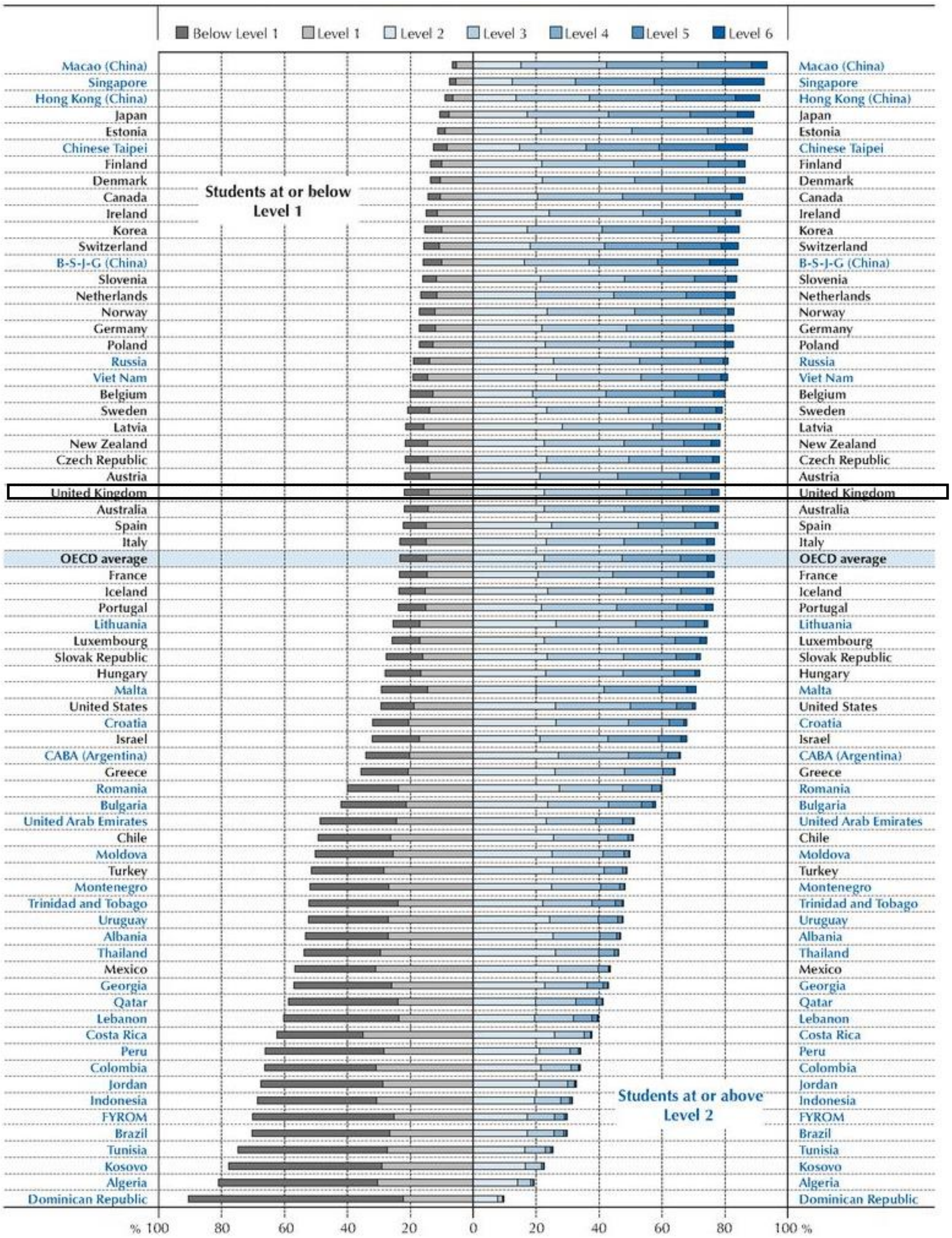


Figure 3 – Student proficiency in mathematics, OECD, 2015, PISA Results (Volume I) Excellence and Equity in Education, Figure 1.5.8, p192. United Kingdom is highlighted.

IMPLICATIONS

Implications of the studies included in the present thesis are discussed within each respective chapter. This section discusses the overarching implications of the thesis in terms of: 1) implications for teachers and school leaders; 2) implications for education policy-makers; and 3) implications for parents.

Implications for teachers and school leaders

Teachers, who are surrounded by hundreds of students every day, are best placed to appreciate that children differ. Yet, there is currently a substantial gap in teacher knowledge between the factors *thought* to explain these differences, and factors that *actually* explain these differences. Chapters 2 and 3 investigated two school environments thought to influence students' academic achievement: school type (selective and non-selective secondary schools) and school quality (as measured by an independent rating body – Ofsted). This thesis found that once pupil characteristics were considered, these school-wide environments explained less than 1% of the variance in achievement at age 16 respectively. In contrast, genetics, which explains up to 60% of the variance in achievement at this age, is often overlooked or misunderstood. Two studies have investigated teachers' perceptions of the heritability of academic achievement (Crosswaite & Asbury, 2018; Walker & Plomin, 2005). Despite being published 13 years apart, they reveal a similar story. Both find that teachers are open to the idea of genetics as a cause of differences in achievement, and that teachers are fairly accurate at estimating heritability. However, Crosswaite et al (2018) found that when teachers were tested on the top ten replicated findings from behavioural genetics (Plomin et al., 2016) (e.g. 'variance in intelligence is the result of many genes' or 'no psychological traits are 100% heritable'), teachers performed worse than chance. This suggests that, while teachers are open to the idea of genetic influence, findings from behavioural genetic research are not well understood.

One barrier to behavioural genetic research being adopted and incorporated into education is its practical application. A question I have been asked numerous times by teachers is: 'What can I actually take away from this research and use tomorrow in my classroom?' This is a difficult question to answer for many reasons, however two reasons stand out. Firstly, twin studies give an estimate of heritability (to what extent differences in a trait are influenced by inherited DNA differences). Heritability is a population statistic, and therefore cannot tell us anything about a specific individual; rather, it tells us about the causes of differences between people at a certain time, in a certain population, measured in a certain way. Heritability tells us about ‘what is’, not ‘what could be’. Therefore, although twin studies show us that, for most traits, genetics plays a substantial role in the differences between people, they cannot tell us anything about the genetic prediction for an individual for a trait of interest. Secondly, although GPS have the advantage of enabling individual-level genetic prediction, they are not currently available for teachers.

GPS present an exciting possibility for teachers who might in the future be able to use them to identify potential problems early on in learning and provide tailored interventions to prevent, rather than treating to cure. However, like all predictors used within education (for example, a student’s prior achievement, whether they have special educational needs, or whether they are entitled to free school meals), they will never be 100% predictive. A GPS created using summary statistics from the 2018 GWA study for years of education with 1.1 million people has been shown to predict up to 11% of the differences in its target trait (years of education) (Lee et al., 2018). Furthermore, it can predict up to 16% in educational achievement (Allegrini et al., in press). This is remarkable progress given that it was only in 2013, that a GPS derived from the first *EduYears* GWA study using 150,000 people explained 2% of the variance in attainment in an independent sample (Rietveld et al., 2013). However, 15% of variance explained still leaves much of the variance unexplained. Therefore, when education does adopt genetics, it will be important to use all data available to teachers, not solely GPS, in order to make accurate predictions to support children in reaching their potential. If GPS are made available to teachers then perhaps there will be a stronger incentive for teachers to better understand genetic research.

Despite these limitations, I believe that teachers and school leaders can take away a couple of points from behavioural genetic research as it stands right now, or at least change the way they approach teaching: 1) the fallacy of group differences; and 2) the environment does not work the way we thought it did.

The fallacy of group differences

Chapters 3 and 4 found that two school-level environments – school type and school quality – made very small unique contributions to children’s achievement, less than 1% in both cases. Many will find this surprising given the frequent media sensationalism regarding selective schooling or Ofsted. However, it highlights a very important issue: average differences mask individual differences. Average differences between schools in terms of their ranking in league tables or published Ofsted ratings can be misunderstood as reflecting each pupil they represent, when in reality there is considerable overlap between students of different schools on a number of outcomes. To illustrate this, Chapter 2 found that grammar school students had higher *EduYears* GPS scores on average compared to those attending state non-selective schools (d = 0.41). However, a Cohen's d of 0.41 is associated with a distribution overlap of approximately 83% between the two distributions. This means that there will be those at grammar schools with low *EduYears* GPS scores, and there will be those at state non-selective schools with high GPS scores. In putting students into groups, it can be easy to forget that they are also individuals.

The environment does not work the way we thought it did

As discussed above, genetics accounts for approximately 40-60% of the differences in achievement throughout schooling. In terms of the environment, one striking finding is that beyond compulsory education, the portion of the variance in education-related traits accounted for by the shared environment (denoted by 'C' in twin studies) appears to decrease (Rimfeld et al, 2016). Why is this important for teachers to know? Because it suggests that the non-shared environment (denoted by 'E') explains more of the differences in student achievement as students get older. It may appear strange that students who, in many ways, share a similar environment go on to achieve very different grades. How can that be? Students learn and experience the world around them based on different prior knowledge, and – partly due to genetics – they differ in their likes, dislikes and their attention (Asbury & Plomin, 2013). A teacher can teach the same lesson (in the same type of school, with the same Ofsted rating) and, while one student will be taking in every word and idea, another student may be thinking about what they are having for lunch. The students have a common environment but often, what is driving the differences between the students are non-shared features of this environment, and how they are engaging with it.

Implications for education policy

Education policy has been reluctant to acknowledge the influence of genetics on education-related traits. The last time genetics was mentioned in relation to its influence on outcomes was by Michael Gove (then UK Education Secretary) in a speech on the future of education given to the Association of School and College Leaders in 2012. He said "…genes do not immutably dictate our destiny - it is the interplay between what we inherit and the environment and culture in which we grow up which determines what we become." Since then, the environment has received much more attention than genetics in terms of education policy debates. However, I believe that there are three things that education policy-makers can take away from behavioural genetics research: 1) support for personalised education and; 2) using heritability as a measure of equal environmental opportunity; and 3) not to be afraid of variation.

Support for personalised education - Learning what you enjoy and enjoying what you learn

Behavioural genetic research has shown that ‘environments’ often show just as much genetic influence as behavioural traits themselves (Plomin & Bergeman, 1991). This thesis shows that school environments are no different; genetics influences the type of school a child attends (Chapter 2), university quality (Chapter 4), and even academic choices, such as choosing to study at university (Chapter 4). This is in line with a vast amount of behavioural genetic literature showing that environments are not ‘out there’ and they do not ‘happen to us’, but instead we select, modify and create our environments, in part based on our genotype. This concept is referred to as gene-environment correlation (Jaffee & Price, 2007, 2008; Knopik et al., 2017). I believe that gene-environment correlation is one of the most powerful discoveries from the behavioural genetic literature, and one that could have implications for educational policy.

An awareness of gene-environment correlation seems to promote a broad curriculum of opportunity – letting children explore their interests, abilities, likes and dislikes, in order to find out what they enjoy. The subjects children are interested in at school tend to be what they are good at (a relationship which appears to go both ways) (Köller, Baumert, & Schnabel, 2001; Luo, Kovas, Haworth, & Plomin, 2011). If, in addition to giving children a basic level of knowledge in key subjects, we offered more opportunities for children to try different things and more freedom for them to choose environments that suit them, we could potentially foster a more positive school experience that encourages children to be active participants in their learning experience (Asbury & Plomin, 2013).

The idea of a more personalised education system, in which children are offered greater choice in the classroom, need not be resource or time-heavy. In a proposal for how education could adopt a more personalised approach, Asbury & Plomin (2015) discuss the promise of new technology as a means of personalising education. Using technology to teach and learn is a field in its infancy but one that appears to be growing

(Hwang & Wu, 2012; Prensky, 2003). Educational computer games are available for maths (McLaren, Adams, Mayer, & Forlizzi, 2018; Van Eck & Dempsey, 2002), history (Squire & Barab, 2004) and computer science (Papastergiou, 2009). Some studies report these games as being at least as effective as traditional teaching. For example, a study looking at the effectiveness of a computer-based maths game found it to be more effective than traditional teaching of the same subject both directly following the experiment ($d = .43$) and one week later ($d = .37$) (McLaren et al., 2018). Further research is needed to understand whether such games could support a more personalised education.

Using heritability as a measure of equal environmental opportunity

Another important finding to come out of behavioural genetic research is that heritability estimates are inextricably linked to the environment. In other words, if the environment changes, then so too can heritability estimates. Although heritability estimates are high across developed countries such as the UK, Australia and the Netherlands (Baker, Treloar, Reynolds, Heath, & Martin, 1996; Bartels, Rietveld, Van Baal, & Boomsma, 2002; Kovas, Haworth, Dale, & Robert, 2007; Krapohl et al., 2014), estimates also vary. A possible explanation of this could be higher heritability estimates in countries with a national curriculum, such as the UK. Why might this be? The answer is that if you reduce variation attributable to the environment (e.g. by giving everyone access to the same curriculum), then a larger proportion of the remaining variation can be explained by other factors, such as genetics. In Chapter 5, we showed that the heritability for university achievement was slightly lower than that for university entrance exams. One possible explanation for this could be the lack of a standardised curriculum between universities. Therefore, the variation in university grades may reflect greater differences between students in terms of how they are taught or how they are examined and may have less to do with genetic differences.

It is important for education policy-makers to understand the push-pull relationship between genetic and environmental influence on education-related traits and use it to their advantage. For example, heritability estimates can be used as a measure of equality of educational provision. The more that differences between individuals are due to inherited DNA differences, the less they will be due to stark differences in environment. A good example of this is a recent paper looking at genetic prediction of educational outcomes in Estonia pre and post-Soviet Union occupation in World War II (Rimfeld et al., 2018). They found that a more egalitarian society following the occupation led to higher heritability estimates compared to during the Soviet era. After the Soviet occupation, receiving an education was less about your rank in society, and more about your natural ability. This has implications for education policy-makers who could use heritability as a tool to measure this key feature of the environment: equality of educational provision.

Do not be afraid of variation

Variation sometimes gets a bad name in education. Variation in a population means that there will always be a low-ability tail and underperforming children, because being 'below average' describes 50% of children if ability is normally distributed. A humorous illustration of this misunderstanding surrounding variation is Garrison Keillor's fictional town of Lake Wobegon (Keillor, 1974-2016), in which 'all the children are above average' or the subtitle of a Schools Week article: 'More than half of multi-academy trusts fell below average for Progress 8 performance" (Staufenberg, 2018). It highlights the problem that society has with the use of statistics. However, once people appreciate that there will always be variation within a population – some of which is driven by genetics – more considered debates can then focus on what goals and strategies should be adopted. For example, does society aim to increase the mean of student grades, reduce the variation, or provide more support for the low-ability tail? These will necessarily involve value judgements, which are outside the scope of this thesis.

Implications for parents

All parents want the best for their children, be that in education, the workplace or in their meaningful relationships. Therefore, finding that school type or Ofsted-rated school quality do not explain substantial variance in children's achievement may initially seem surprising or even disappointing to some parents. This finding may be especially disappointing to parents who have put substantial time and resources into getting their child into a selective school or moving to be closer to an Ofsted-rated 'Outstanding' secondary school. However, it highlights a well-replicated behavioural genetic finding that, for many educationally-relevant traits, the common environment, such as general features of schools, explain little variance, especially as children get older. The implication here for parents is that focusing exclusively on school environments, such as school quality or school type, as a determinant of academic achievement is more likely a red herring than a silver bullet. Parents should not be led to believe that merely getting their child into a selective school or an Ofsted-rated 'Outstanding' quality school has any academic guarantees. It can be easy to over-estimate the influence of general features of the school environment, and under-estimate genetic influence. Of course, there are many reasons, beside academic achievement, why parents may opt to send their children to a particular type of school.

FUTURE DIRECTIONS

We have entered an exciting time in terms of genetic prediction of behavioural traits. Using genome-wide polygenic scores, it is possible to predict up to 16% of the variance in achievement (Allegrini et al., in press). However, although this research is steaming ahead, there has been little work on its translation to key stakeholders, such as policy-makers, teachers, parents and students. This is despite the fact that there is now direct-to-consumer genetic testing, for example 23 and me (https://www.23andme.com/en-gb/), as well as services that can create polygenic scores for individuals for a number of traits, for example DNA Land (https://dna.land/). It is not unrealistic to imagine that in the future, parents may opt to get polygenic scores for their children and wish to hold schools to account in supporting their unique genetic needs. Therefore, research is needed to: 1) understand the current level of public awareness of polygenic scores, with emphasis on teacher knowledge; and 2) explore the most effective way of communicating polygenic score results to parents, students and teachers.

Public understanding of polygenic scores

Previous research has investigated public understanding of genetic influence (heritability estimates) on different traits in the UK (Crosswaite & Asbury, 2018; Walker & Plomin, 2005), the US (Willoughby et al., 2018) and Russia (Chapman et al., 2017). Two of these studies focused specifically on teacher's perceptions of genetic influence (Crosswaite & Asbury, 2018; Walker & Plomin, 2005). However, so far, there has not been any research investigating public understanding of genome-wide polygenic scores. Part of this is due to the infancy of the field. Research using polygenic scores has not been around long, and has only started to explain substantial variance in education-related traits in the last few years (Plomin & von Stumm, 2018). However, I believe that this is the perfect time for canvassing the views of the public regarding polygenic scores. It will be important to have an emphasis on teacher understanding, to get a baseline in which to measure progress of public understanding in the years to come and as polygenic scores become more predictive of educationally-relevant traits.

Communicating polygenic score results

When polygenic scores do become more mainstream, it will be vital that the results are clearly and accurately communicated to avoid misconceptions. Previous research has investigated the impact of teaching university students about their personal genome, including student attitudes towards genome sequencing, decision-making (whether to get their genome sequenced or not), psychological wellbeing, genomics knowledge and engagement (Linderman et al., 2018). However, to my knowledge, there has not been any research into the impact of communicating personal polygenic scores. One research project in the US 'Spit for Science' is currently collecting genetic and phenotypic data on university students to investigate genetic and environmental influences on substance use and emotional health (Dick et al., 2014). However, while they plan on creating polygenic scores, there is no information on whether they intend to communicate their findings to the participants. Understanding the meaning behind polygenic scores for a particular trait not only requires some level of genetics knowledge, but also statistical knowledge, such as an understanding of variance, probability and effect size. Better understanding of how to communicate polygenic scores in a meaningful and sensitive way will become increasingly valuable as polygenic scores become more mainstream. Crucially, it will be necessary to assess whether communicating polygenic scores is a useful exercise and whether it helps children's learning and development.

As well as future research into the communication of polygenic scores, I also hope to extend research into the impact of school environments on educational achievement and other academically-relevant traits. In particular, I plan to conduct a follow-up study looking at the impact of different school types beyond academic achievement at age 16, by extending this research to look at the impact of school type on university admissions, university success and career choice and prospects. Although there has been previous research investigating the influence of school types on degree grade (Higher Education Funding Council for England, 2013) earnings (Broughton, Ezeyi, Hupkau, Keohane, & Shorthouse, 2014) and career choice (Sutton Trust, 2009), these studies have not always taken student covariates into account, for example general cognitive ability, prior achievement, family socioeconomic status or genetics, which may be driving the majority of the association between school type and future success. It will be important to consider these factors when assessing the influence of school environments on future success.

Conclusion

To conclude, the present thesis investigated the genetic and environmental influences on achievement in secondary school and beyond, with a focus on student covariates and school-wide factors. This thesis illustrated the importance of genetics throughout education, highlighted the small effect of school-wide factors on individual differences in achievement and discussed the emergence of polygenic scores. The findings of this thesis, and behavioural genetic research more generally, support the trend towards personalised education, in which teaching is tailored to the specific profile of each student, and students are active participants in finding and forming their learning environments. I finish by going back to the quote presented at the beginning of this thesis: ‘The mind is not a vessel to be filled, but a fire to be kindled’ (Plutarch, in *Moralia* 48C). Indeed, education should spark the flames of children’s’ life-long interests.

References

Allegrini, A. G., Selzam, S., Rimfeld, K., Von Stumm, S., Pingault, J.-B., & Plomin, R. (in press). Genomic prediction of cognitive traits in childhood and adolescence. *bioRxiv*.

Asbury, K., & Plomin, R. (2013). *G is for genes: The impact of genetics on education and achievement* (Vol. 24): John Wiley & Sons.

Baker, L. A., Treloar, S. A., Reynolds, C. A., Heath, A. C., & Martin, N. G. (1996). Genetics of educational attainment in Australian twins: sex differences and secular changes. *Behavior genetics, 26*(2), 89-102.

Bartels, M., Rietveld, M. J., Van Baal, G. C. M., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research and Human Genetics, 5*(6), 544-553.

Broughton, N., Ezeyi, O., Hupkau, C., Keohane, N., & Shorthouse, R. (2014). *An independent evaluation*. London, UK: The Social Market Foundation.

Chapman, R., Likhanov, M., Selita, F., Zakharov, I., Smith-Woolley, E., & Kovas, Y. (2017). *Genetic Literacy And Attitudes Survey (Iglas): International Population-Wide Assessment.* Paper presented at the The European Proceedings of Social & Behavioural Sciences EPSBS.

Crosswaite, M., & Asbury, K. (2018). Teacher beliefs about the aetiology of individual differences in cognitive ability, and the relevance of behavioural genetics to education. *British Journal of Educational Psychology, 12224*, 1-16.

Deary, I. J., Yang, J., Davies, G., Harris, S. E., Tenesa, A., Liewald, D., . . . Corley, J. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature, 482*(7384), 212-215.

Dick, D., Nasim, A., Edwards, A. C., Salvatore, J., Cho, S. B., Adkins, A., . . . Clifford, J. (2014). Spit for Science: launching a longitudinal study of genetic and environmental influences on substance use and emotional health at a large US university. *Frontiers in genetics, 5*, 47.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*(3), 385-400.

Higher Education Funding Council for England. (2013). *Higher education and beyond: Outcomes from full-time first degree study*. London, UK: Higher Education Funding Council for England.

Hwang, G. J., & Wu, P. H. (2012). Advancements and trends in digital game-based learning research: a review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology, 43*(1), E6-E10.

Jaffee, S. R., & Price, T. S. (2007). Gene–environment correlations: A review of the evidence and implications for prevention of mental illness. *12*(5), 432-442.

Jaffee, S. R., & Price, T. S. (2008). Genotype–environment correlations: implications for determining the relationship between environmental exposures and psychiatric illness. *Molecular Psychiatry, 7*(12), 496-499.

Keillor, G. (1974-2016). A Prairie Home Companion. Minnesota, US: Minnesota Public Radio.

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Archives of general psychiatry, 60*(9), 929-937.

Knopik, V. S., Neiderheiser, J., DeFries, J. C., & Plomin, R. (2017). *Behavioral genetics* (7th Ed.). New York: Worth.

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, 448-470.

Kovas, Y., Haworth, C., Dale, P. S., & Robert, P. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for research in Child Development, 72*(3), vii, 1-144.

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., . . . Dale, P. S. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences, 111*(42), 15273-15278.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., . . . Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics, 50*(1), 1112-1121.

Linderman, M. D., Sanderson, S. C., Bashir, A., Diaz, G. A., Kasarskis, A., Zinberg, R., . . . Schadt, E. E. (2018). Impacts of incorporating personal genome sequencing into graduate genomics education: a longitudinal study over three course years. *BMC medical genomics, 11*(1), 5.

Luo, Y. L., Kovas, Y., Haworth, C. M., & Plomin, R. (2011). The etiology of mathematical self-evaluation and mathematics achievement: Understanding the relationship using a cross-lagged twin study from ages 9 to 12. *Learning and individual differences, 21*(6), 710-718.

McLaren, B. M., Adams, D. M., Mayer, R. E., & Forlizzi, J. (2018). A computer-based game that promotes mathematics learning more than a conventional approach

Gamification in Education: Breakthroughs in Research and Practice (pp. 415-437): IGI Global.

Ndaji, F., Little, J., & Coe, R. (2016). *A comparison of Academic Achievement in Independent and State Schools*. Durham, UK: Centre for Evaluation and Monitoring, Durham University.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European journal of social psychology, 15*(3), 263-280.

Organisation for Economic Co-operation and Development. (2017). *PISA 2015 Results (Volume I) Excellence and Equity in Education*: OECD Publishing.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education, 52*(1), 1-12.

Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influence on "environmental" measures. *Behavioral and Brain Sciences, 14*(3), 373-386.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science, 11*(1), 3-23.

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*(3), 148-159.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology, 88*(5), 879-903.

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE), 1*(1), 21-21.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., . . . D, K. P. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *science, 340*(6139), 1467-1471.

Rimfeld, K., Krapohl, E., Trzaskowski, M., Coleman, J. R., Selzam, S., Dale, P. S., . . . Plomin, R. (2018). Genetic influence on social outcomes during and after the Soviet era in Estonia. *Nature human behaviour*, 1.

Roorda, D. L., Koomen, H. M., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of educational research, 81*(4), 493-529.

Squire, K., & Barab, S. (2004). *Replaying history: Engaging urban underserved students in learning world history through computer simulation games.* Paper presented at the Proceedings of the 6th international conference on Learning sciences.

Staufenberg, J. (2018). MAT league tables 2017: The best and worst performers, and 4 other findings.

Sullivan, A., & Heath, A. (2002). *State and Private Schools in England and Wales*. Oxford, UK: Nuffield College, Oxford and Department of Sociology, University of Oxford.

Sutton Trust. (2009). *The Educational Backgrounds of Leading Lawyers, Journalists, Vice Chancellors, Politicians, Medics and Chief Executives*. London, UK: The Sutton Trust.

Van Eck, R., & Dempsey, J. (2002). The effect of competition and contextualized advisement on the transfer of mathematics skills a computer-based instructional simulation game. *Educational Technology Research and Development, 50*(3), 23-41.

Verhulst, F. C., & Van der Ende, J. (1992). Agreement between parents' reports and adolescents' self-reports of problem behavior. *Journal of Child Psychology and Psychiatry, 33*(6), 1011-1023.

Walker, S. O., & Plomin, R. (2005). The nature–nurture question: Teachers' perceptions of how genes and the environment influence educationally relevant behaviour. *Educational Psychology, 25*(5), 509-516.

Willoughby, E. A., Love, A. C., McGue, M., Iacono, W. G., Quigley, J., & , & Lee, J. J. (2018). Free will, determinism, and intuitive judgments about the heritability of behavior. *OSF Preprints*. doi: 10.31219/osf.io/ezg2j

Appendices

The supplementary materials for chapters 2-5, as references in the text, are attached as appendices

Appendix 1: Supplementary materials for Chapter 2
Appendix 2: Supplementary materials for Chapter 3
Appendix 3: Supplementary materials for Chapter 4
Appendix 4: Supplementary materials for Chapter 5

Appendix 1: Supplementary materials for Chapter 2

Polygenic score for educational attainment captures DNA variants shared between personality-related traits and educational achievement

Smith-Woolley, E., Selzam, S., & Plomin, R.

Supplementary Methods

Methods S1 – Genotyping and quality control

Methods S2 – Structural equation model of non-cognitive domains, GCSE results and polygenic scores

Supplementary Tables

Table S1 – Descriptive statistics of all variables

Table S2 – Sensitivity analysis of missingness of personality/motivation composites on socio-economic status, general cognitive ability, and GCSE grades

Table S3 – Numbers of SNPs included in EduYears GPS, Neuroticism GPS and Wellbeing GPS

Supplementary Figures

Figure S1 – Correlations across all individual measures of personality and motivation, the personality/motivation composites and polygenic scores

Figure S2 – Correlations across all polygenic scores and personality/motivation domains

Figure S3 – Correlations across all EduYears GPS thresholds and personality/motivation domains

Figure S4 – Correlations across all Neuroticism GPS thresholds and personality/motivation domains

Figure S5 – Correlations across all Wellbeing GPS thresholds and personality/motivation domains

Supplementary Methods

Methods S1 – Genotyping and quality control

DNA for 4,649 individuals was extracted from saliva and buccal cheek swab samples and hybridised to HumanOmniExpressExome-8v1.2 genotyping arrays at the Institute of Psychiatry, Psychology and Neuroscience Genomics & Biomarker Core Facility. The raw image data from the array were normalised, pre-processed, and filtered in GenomeStudio according to Illumina Exome Chip SOP v1.4. (http://confluence.brc.iop.kcl.ac.uk:8090/display/PUB/Production+Version%3A+Illumina+Exome+Chip+SOP+v1.4). In addition, prior to genotype calling, 869 multi-mapping SNPs and 353 samples with callrate < 0.95 were removed. The ZCALL program (see Web resources section) was used to augment the genotype calling for samples and SNPs that passed the initial QC.

DNA from 3,665 samples was extracted from buccal cheek swabs and genotyped at Affymetrix, Santa Clara, California, USA. Samples were successfully hybridised to AffymetrixGeneChip 6.0 SNP genotyping arrays (http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf) using experimental protocols recommended by the manufacturer (Affymetrix Inc., Santa Clara, CA). The raw image data from the arrays were normalised and pre-processed at the Wellcome Trust Sanger Institute, Hinxton, UK for genotyping as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/) according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Genotypes for the Affymetrix arrays were called using CHIAMO (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html).

After initial quality control and genotype calling, the same quality control was performed on the samples genotyped on the Illumina and Affymetrix platforms separately using PLINK (Chang et al., 2015; Purcell et al., 2007), R (R Core Team, 2015), and vcftools (Danecek et al., 2011).

Samples were removed from subsequent analyses on the basis of call rate (<0.99), suspected non-European ancestry, heterozygosity, array signal intensity, and relatedness. SNPs were excluded if the minor allele frequency was <0.5%, if more than 1% of genotype data were missing, or if the Hardy Weinberg *p*-value was lower than 10^{-5}. Non-autosomal markers and indels were removed. Association between the SNP and the platform, batch, or plate on which samples were genotyped was calculated;

SNPs with an effect *p*-value less than 10^{-3} were excluded. A total sample of 6,710 samples, with 3,617 individuals and 600,034 SNPs genotyped on Illumina and 3,093 individuals and 525,859 SNPs genotyped on Affymetrix remained after quality control.

Genotypes from the two platforms were separately imputed using the Haplotype Reference Consortium (McCarthy et al., 2016) and Minimac3 1.0.13 (Fuchsberger, Abecasis, & Hinds, 2014; Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012) available on the *Michigan Imputation Server* as reference data. A series of quality checks was performed before merging data from the two platforms' imputation (e.g. platform effects, allele frequencies by imputation quality). For the present analyses, we limited our analyses to variants genotyped or imputed at info >.70 on both platforms, allele frequency difference between platforms smaller than 5%, and Hardy Weinberg *p*-value was greater than 10^{-5}. Using these criteria, 7,581,516 genotyped and well-imputed SNPs were retained for the analyses.

We performed principal component analysis on a subset of 42,859 common (MAF>5%) autosomal HapMap3 SNPs (Haplotype Reference Consortium, 2016), after stringent pruning to remove markers in linkage disequilibrium ($r^2 > 0.1$) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected.

References:

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience, 4*(1), 7.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

Fuchsberger, C., Abecasis, G. R., & Hinds, D. A. (2014). minimac2: faster genotype imputation. *Bioinformatics, 31*(5), 782-784.

Haplotype Reference Consortium. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics, 48*(10), 1279-1283.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics, 44*(8), 955-959.

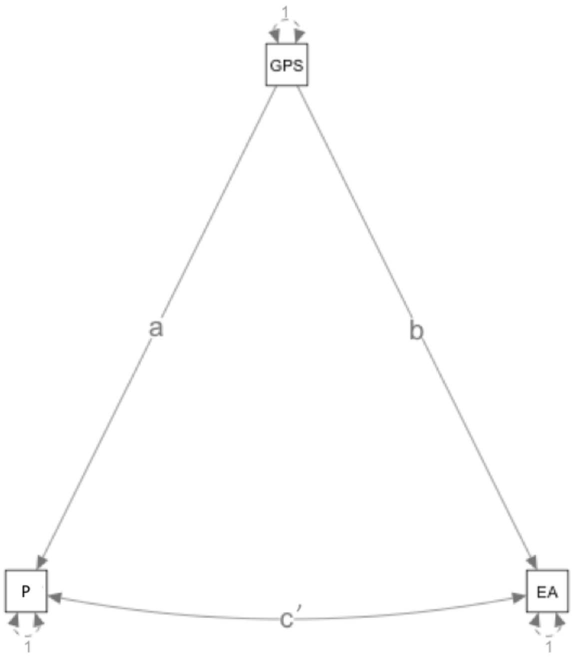
McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., . . . Sharp, K. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics, 48*(10), 1279.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

R Core Team. (2015). R: A language and environment for statistical computing.

Methods S2 – Structural equation model of non-cognitive domains, GCSE results and polygenic scores

To test the extent to which the covariance between non-cognitive domains and GCSE results are explained by the polygenic scores, we used structural equation modelling. Because we assume causality from polygenic score (an aggregate score of DNA variants) to outcome variables, we applied the following model to our z-standardised variables:



Note: P = personality trait, EA = educational achievement defined as GCSE results, GPS = genome-wide polygenic score

Paths a, b and c are the beta effect size parameters. The GPS effect is described by the product of a and b, which is the pathway from the causal variable GPS to P, and GPS to EA. Path c' describes the residual relationship between P and EA after accounting for the effects of the causal variable GPS in P and EA, respectively. The total effect can be derived by summing the effects of the residual and the indirect path, described as c' + ab. To calculate the proportion of the total effect that is explained by the causal variable GPS, the effect of the indirect path ab is divided by the total effect c' + ab.

Supplementary Tables

Table S1 – Descriptive statistics of all variables

		Mean (SD)			ANOVA of gender		ANOVA of age	
	N	Whole sample	Males	Females	*F*	R^2	*F*	R^2
Motivation Composite	**2032**	**0.00 (1.00)**	**0.00 (0.98)**	**0.00 (1.02)**	-	-	-	-
PISA math self-efficacy	2031	17.58 (5.47)	19.09 (4.87)	16.53 (5.62)	119.56**[1]	0.05	0.01	<0.01
PISA math interest	2032	2.54 (0.93)	2.65 (0.90)	2.46 (0.94)	22.47**[1]	0.01	0.11	<0.01
PISA time spent on math	2014	4.47 (1.71)	4.52 (1.82)	4.43 (1.64)	1.07[1]	<0.01	0.02	<0.01
Academic self-concept	1880	3.55 (0.63)	3.68 (0.61)	3.46 (0.62)	56.80**	0.03	0.02	<0.01
Attitude towards key subjects	2024	2.54 (0.49)	2.51 (0.50)	2.55 (0.49)	3.44	<0.01	1.07	<0.01
School engagement	2027	3.01 (0.68)	3.00 (0.68)	3.01 (0.67)	0.20	<0.01	0.00	<0.01
Openness Composite	**1882**	**0.00 (1.00)**	**0.00 (0.96)**	**0.00 (1.02)**	-	-	-	-
Extraversion	1858	3.68 (0.61)	3.65 (0.61)	3.69 (0.61)	1.97	<0.01	1.74	<0.01
Openness	1857	3.59 (0.58)	3.58 (0.58)	3.60 (0.58)	0.84	<0.01	0.39	<0.01
Ambition	1880	3.90 (0.68)	3.93 (0.64)	3.87 (0.70)	3.29[1]	<0.01	0.06	<0.01
Curiosity	2132	4.78 (0.91)	4.93 (0.90)	4.67 (0.90)	44.45**	0.02	2.09	<0.01
Hopefulness	2137	4.70 (0.72)	4.81 (0.69)	4.61 (0.73)	40.18**	0.02	2.99	<0.01
Conscientiousness Composite	**1793**	**0.00 (1.00)**	**0.00 (0.97)**	**0.00 (1.02)**	-	-	-	-
Conscientiousness	1854	3.72 (0.60)	3.65 (0.60)	3.77 (0.60)	19.71**	0.01	0.13	<0.01
SDQ Conduct scale (r.)	3830	8.40 (1.43)	8.34 (1.43)	8.45 (1.43)	5.44*	<0.01	8.46**	<0.01
SDQ hyperactivity (r.)	3830	6.51 (2.28)	6.44 (0.92)	6.56 (2.27)	2.75	<0.01	3.16	<0.01
GRIT	1909	3.28 (0.60)	3.24 (0.59)	3.31 (0.60)	7.72**	<0.01	2.51	<0.01
SWAN Inattention	865	4.80 (0.99)	4.79 (0.92)	4.80 (1.04)	0.01[1]	<0.01	3.81	<0.01
SWAN Hyperactivity	866	4.69 (0.85)	4.70 (0.83)	4.68 (0.85)	0.08	<0.01	0.82	<0.01
Agreeableness Composite	**2013**	**0.00 (1.00)**	**0.02 (0.99)**	**-0.02 (1.01)**	-	-	-	-
Agreeableness	1853	3.67 (0.56)	3.54 (0.55)	3.76 (0.56)	68.68**	0.04	0.06	<0.01
Prosocial behaviour	3830	7.17 (1.94)	6.56 (1.96)	7.65 (1.79)	313.64**[1]	0.08	10.63**	<0.01
Gratitude	2139	5.80 (0.86)	5.69 (0.85)	5.88 (0.86)	25.17**	0.01	0.3	<0.01
Neuroticism Composite	**3831**	**0.00 (1.00)**	**0.01 (0.89)**	**0.00 (1.08)**	-	-	-	-
Neuroticism (r.)	1858	3.42 (0.68)	3.56 (0.64)	3.33 (0.69)	50.83**	0.03	5.48*	<0.01

Cognitive Disorganisation	3830	7.10 (2.85)	7.71 (2.70)	6.63 (2.88)	142.46**[1]	0.09	3.82	<0.01
CASI anxiety (r.)	3831	28.15 (5.86)	30.11 (4.64)	26.61 (6.24)	397.13**[1]	0.04	0.99	<0.01
MFQ (r.)	3832	1.72 (0.34)	1.80 (0.26)	1.66 (0.38)	180.21**[1]	<0.01	5.76*	<0.01
Life satisfaction	2134	4.61 (0.62)	4.64 (0.57)	4.58 (0.64)	5.36*[1]	<0.01	1.65	<0.01
Subjective happiness	2141	5.22 (1.17)	5.22 (1.14)	5.22 (1.19)	0.00	<0.01	0.31	<0.01
Optimism	1909	3.23 (0.72)	3.31 (0.70)	3.17 (0.73)	17.52**	<0.01	1.74	<0.01
Peer problems (r.)	3830	1.54 (1.51)	1.60 (1.51)	1.50 (1.51)	4.39*	<0.01	0.03	<0.01
GCSE	**5361**	**8.80 (1.27)**	**8.76 (1.26)**	**8.84 (1.26)**	5.54*	<0.01	15.07**	<0.01
General cognitive ability[+]	**2633**	**0.06 (0.81)**	**0.05 (0.82)**	**0.06 (0.81)**	**-**	**-**	**-**	**-**
***EduYears* GPS**	**6710**	**0.00 (1.00)**	**0.04 (1.00)**	**0.02 (1.00)**	-	-	-	-
Neuroticism GPS	**6710**	**0.00 (1.00)**	**0.03 (1.00)**	**0.01 (1.00)**	-	-	-	-
Wellbeing GPS	**4584**	**0.00 (1.00)**	**0.00 (1.00)**	**0.01 (1.02)**	-	-	-	-

Note: Means and standard deviations for individual measures are calculated based on raw data. Means and standard deviations for domains are calculated with z-standardised age and sex regressed data. Values of standard deviation are given in parentheses. (r) = reversed items. [+] = standardization of the individual cognitive scales assessed at age 7, 12 and 16 was required to form this composite. *N*= sample size after exclusions. ANOVA performed on one randomly selected twin per pair to test the effect of sex and age. Results = *F* statistic; [1] = homogeneity of variance was not equal, Welch test used instead * = *p*<.05; ** = p<.01; R^2= proportion of variance explained by sex and age

Table S2 – Sensitivity analysis of missingness of personality/motivation composites on socio-economic status, general cognitive ability, and GCSE grades

	Present *M* (*SD*)	Missing *M* (*SD*)	*t*	*p*	*df*	R^2
Motivation Composite						
SES	0.013	-0.008	0.722	0.470	5125	<0.001
General cognitive ability	0.028	-0.063	2.175	0.030	2631	0.002
GCSE	0.111	-0.072	5.855	5.19×10^{-9}	4204	0.008
Openness Composite						
	Present *M* (*SD*)	Missing *M* (*SD*)	*t*	*p*	*df*	R^2
SES	0.029	-0.016	1.542	0.123	5125	<0.001
General cognitive ability	0.039	-0.076	2.792	0.005	2631	0.003
GCSE	0.139	0.081	7.025	2.57×10^{-12}	4204	0.012
Conscientiousness Composite						
	Present *M* (*SD*)	Missing *M* (*SD*)	*t*	*p*	*df*	R^2
SES	0.027	-0.014	1.375	0.169	5125	<0.001
General cognitive ability	0.076	-0.110	4.728	0.000002	2631	0.008
GCSE	0.118	-0.065	5.713	1.19×10^{-8}	4204	0.008
Agreeableness Composite						
	Present *M* (*SD*)	Missing *M* (*SD*)	*t*	*p*	*df*	R^2
SES	0.023	-0.014	1.281	0.200	5125	<0.001
General cognitive ability	0.028	-0.062	2.143	0.032	2631	0.002
GCSE	0.11	-0.0713	5.772	8.40×10^{-9}	4204	0.008
Neuroticism Composite						
	Present *M* (*SD*)	Missing *M* (*SD*)	*t*	*p*	*df*	R^2
SES	0.051	-0.129	5.819	$6.29x^{-9}$	5125	0.007
General cognitive ability	0.056	-0.17	5.034	$5.14x^{-7}$	2631	0.010
GCSE	0.059	-0.189	6.85	$8.46x^{-12}$	4204	0.011

Note*:* SES = socio-economic status; GCSE = General Certificate for Secondary Education; present = data is non-missing for the respective composite; missing = data is missing for the respective composite.

Table S3 – Numbers of SNPs included in EduYears GPS, Neuroticism GPS and Wellbeing GPS

EduYears		Neuroticism		Wellbeing	
*p*T	*N*SNPs	*p*T	*N*SNPs	*p*T	*N*SNPs
0.001	2,162	0.001	4,841	0.001	1,420
0.05	19,415	0.05	58,410	0.05	19,956
0.1	30,086	0.1	93,493	0.1	30,636
0.2	46,636	0.2	149,091	0.2	47,032
0.3	60,012	0.3	194,300	0.3	58,135
0.4	71,382	0.4	232,481	0.4	69,456
0.5	81,149	0.5	265,272	0.5	77,897

Note: *p*T = GWAS association *p*-value threshold under which GPS was constructed; *N*SNPs = Number of SNPs included in the GPS.

Supplementary Figures

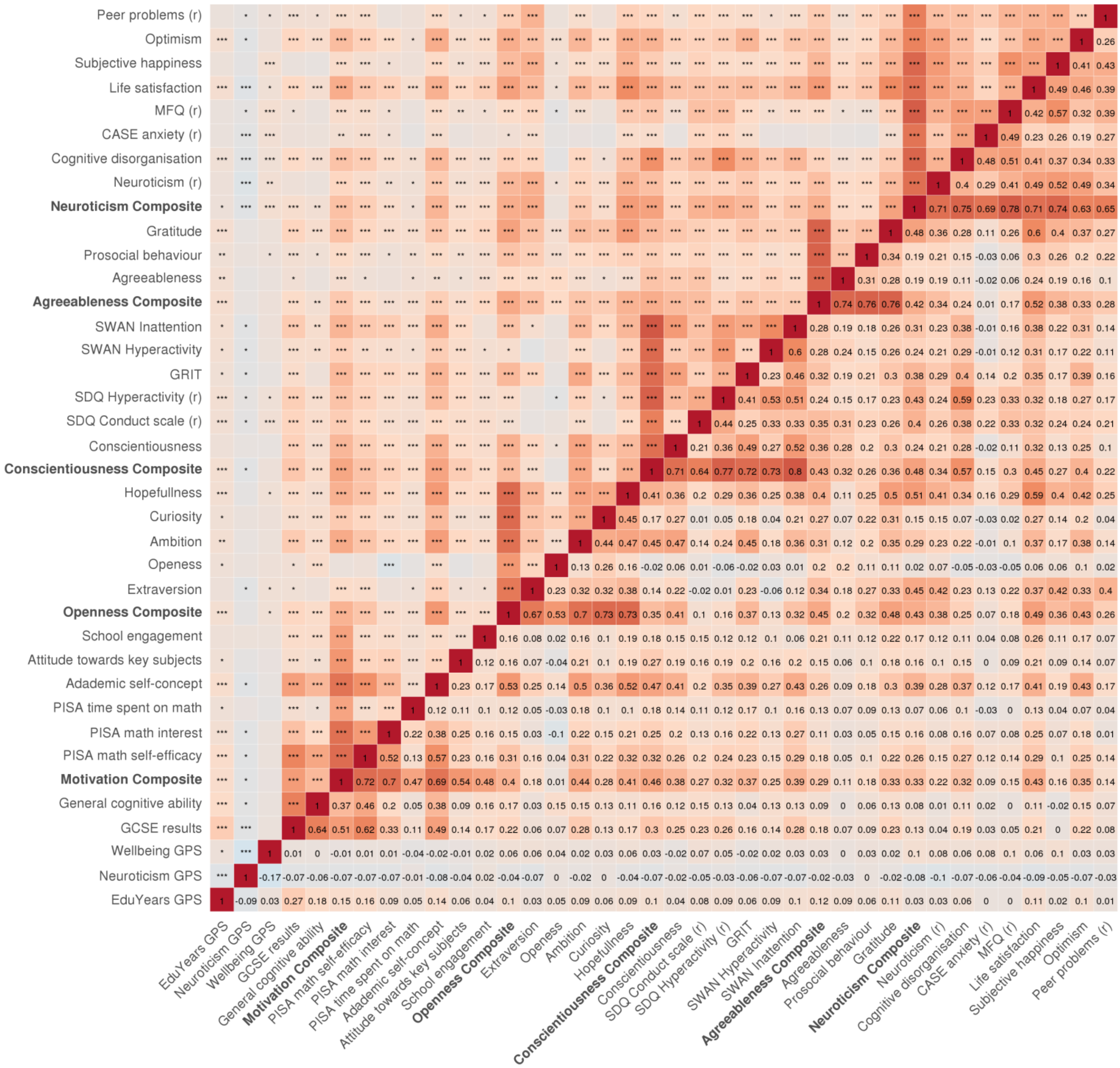


Figure S1 – Correlations across all individual measures of personality and motivation, the personality/motivation composites and polygenic scores. *Note:* (r.) = recoded so that higher scores were positive, i.e. less conduct problems. Variable labels in bold represent composites made up of the succeeding individual scales. * = $p < 0.05$; ** = $p < 0.001$; *** = $p < 0.0001$.

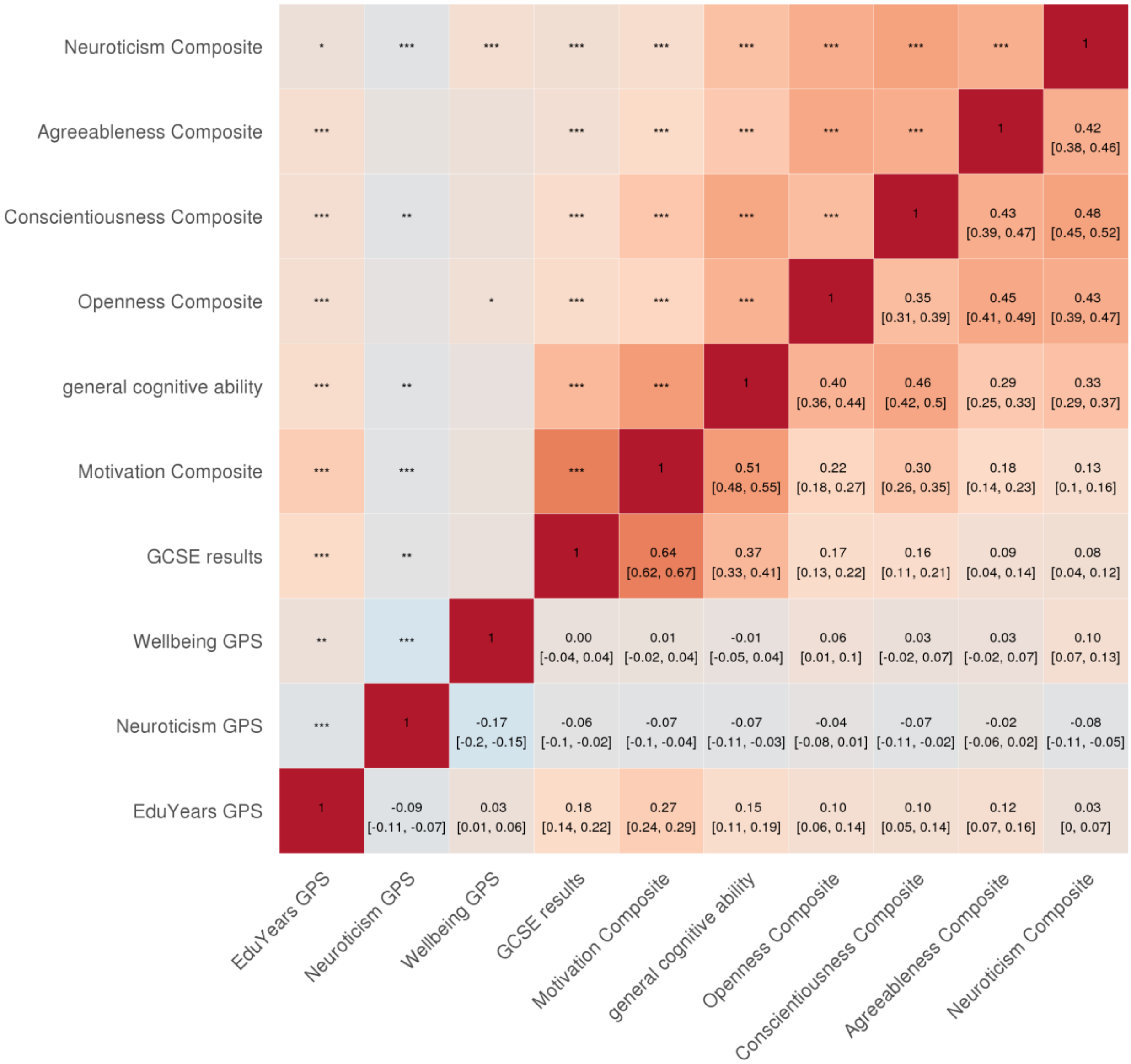


Figure S2 – Correlations across all polygenic scores and personality/motivation domains. Note: 95% confidence intervals of the Pearson's correlation coefficients shown in square brackets. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

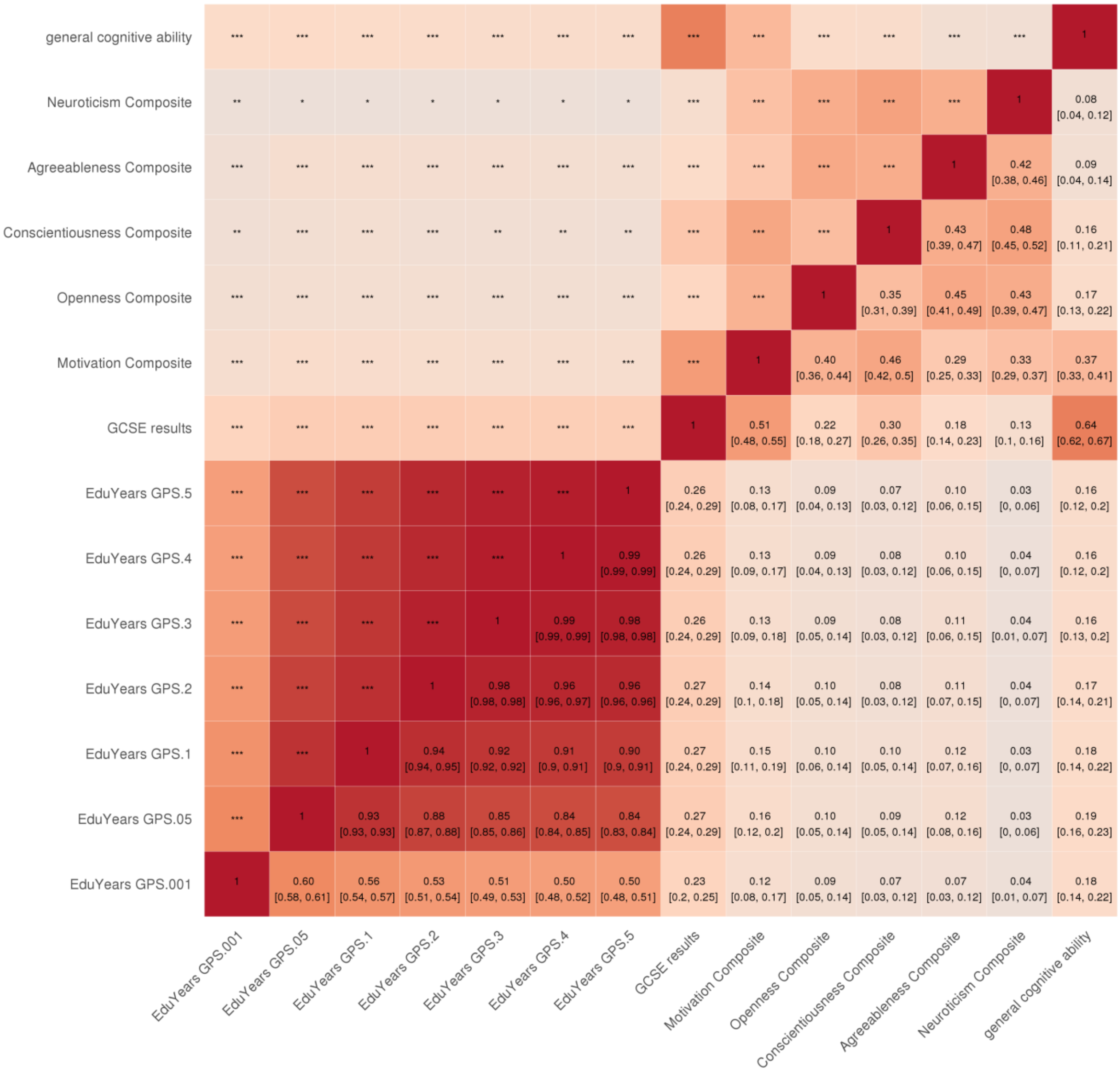


Figure S3 – Correlations across all EduYears GPS thresholds and personality/motivation domains. *Note:* 95% confidence intervals of the Pearson's correlation coefficients shown in square brackets. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

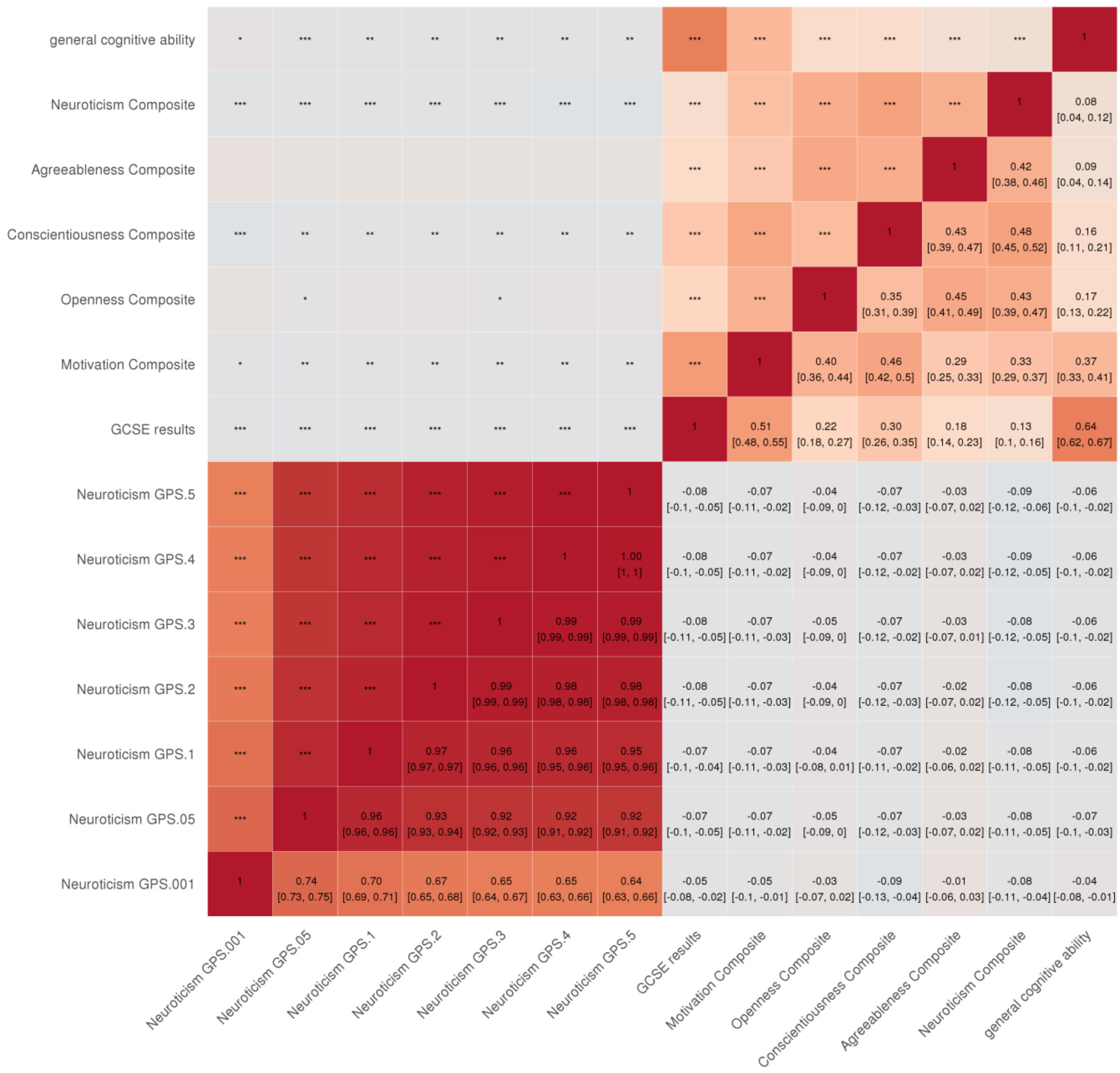


Figure S4 – Correlations across all Neuroticism GPS thresholds and personality/motivation domains. *Note:* 95% confidence intervals of the Pearson's correlation coefficients shown in square brackets. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

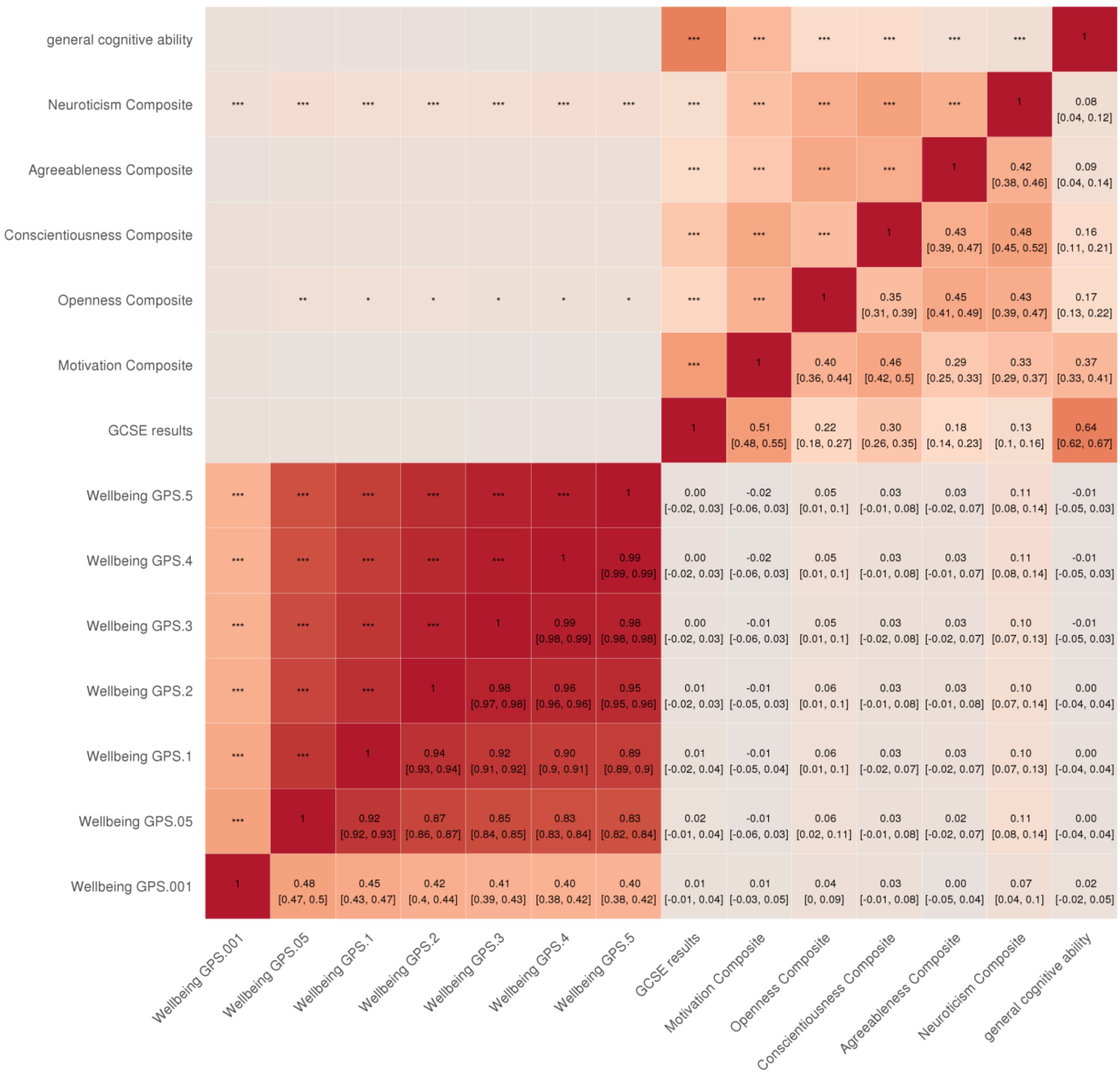


Figure S5 – Correlations across all Wellbeing GPS thresholds and personality/motivation domains. *Note:* 95% confidence intervals of the Pearson's correlation coefficients shown in square brackets. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

Appendix 2: Supplementary Material for Chapter 3

Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them

Smith-Woolley, E., Pingault, J B., Selzam, S., Rimfeld, K., Krapohl, E., von Stumm, S., Asbury, K., Dale, P. S., Young, T., Allen, R., Kovas, Y., & Plomin, R.

Supplementary methods:

Methods S1 – Details on genotyping

Methods S2 – Creating the school type variables

Methods S3 – Hierarchical linear regression to calculate adjusted means for school type

Supplementary Tables:

Table S1 – Analysis of variance (ANOVA) and planned contrasts for *EduYears* GPS between students of three school types: state non-selective, grammar and private schools

Table S2 – Analysis of variance (ANOVA) and planned contrasts for *EduYears* GPS between students of five school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private schools

Table S3 – Correlation matrix

Table S4 – Hierarchical regression analysis of *EduYears* GPS, controlling for selection factors for students of three school types: state non-selective, grammar and private schools

Table S5 – Hierarchical regression analysis of EduYears GPS, controlling for selection factors for students of five school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private schools

Table S6 – Regression analysis of predictors of mean GCSE for three school types: state non-selective, grammar and private schools

Table S7 – Regression analysis of predictors of mean GCSE for three school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private schools

Table S8 – Descriptive statistics

Supplementary Figures:

Figure S1 – *EduYears* GPS plotted means (and 95% confidence intervals) for students of five school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private schools

Figure S2 – *EduYears* GPS plotted means (and 95% confidence intervals) controlling for selection factors between students of 3 school types: non-selective state, grammar and private

Figure S3 – EduYears GPS plotted means (and standard errors) controlling for selection factors between 5 school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private school

Figure S4 – The plotted means (and 95% confidence intervals) for unadjusted GCSE, GCSE controlling for *EduYears* GPS, GCSE controlling for SES, GCSE controlling for prior ability, GCSE controlling for prior achievement and GCSE controlling for all variables between 5 school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private school

Figure S5 – Mean *EduYears* GPS (and 95% confidence intervals) between state non-selective, grammar and private school for several EduYears GPS p-value cut-off

Methods S1 – Details on genotyping

DNA for 4,649 individuals was extracted from saliva and buccal cheek swab samples and hybridised to HumanOmniExpressExome-8v1.2 genotyping arrays at the Institute of Psychiatry, Psychology and Neuroscience Genomics & Biomarker Core Facility. The raw image data from the array were normalised, pre-processed, and filtered in GenomeStudio according to Illumina Exome Chip SOP v1.4. (http://confluence.brc.iop.kcl.ac.uk:8090/display/PUB/Production+Version%3A+Illumina+Exome+Chip+SOP+v1.4). In addition, prior to genotype calling, 869 multi-mapping SNPs and 353 samples with call rate <.95 were removed.

DNA from 3,665 samples was extracted from buccal cheek swabs and genotyped at Affymetrix, Santa Clara, California, USA. Samples were successfully hybridised to AffymetrixGeneChip 6.0 SNP genotyping arrays (http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf) using experimental protocols recommended by the manufacturer (Affymetrix Inc., Santa Clara, CA). The raw image data from the arrays were normalised and pre-processed at the Wellcome Trust Sanger Institute, Hinxton, UK for genotyping as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/) according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Genotypes for the Affymetrix arrays were called using CHIAMO (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html). After initial quality control and genotype calling, the same quality control was performed on the samples genotyped on the Illumina and Affymetrix platforms separately using PLINK (Chang et al., 2015; Purcell et al., 2007), R (R Core Team, 2016), and vcftools (Danecek et al., 2011).

Samples were removed from subsequent analyses on the basis of call rate (<0.99), suspected non-European ancestry, heterozygosity, array signal intensity, and relatedness (IBD cut-off 0.05). SNPs were excluded if the minor allele frequency was <0.5%, if more than 1% of genotype data were missing, or if the Hardy Weinberg *p*-value was lower than 10^{-5}. Non-autosomal markers and indels were removed. Association between the SNP and the platform, batch, or plate on which samples were genotyped was calculated; SNPs with an effect *p*-value less than 10^{-3} were excluded. A total sample of 6,710 samples, with 3,617 individuals and 600,034 SNPs genotyped on

Illumina and 3,093 individuals and 525,859 SNPs genotyped on Affymetrix remained after quality control.

Genotypes from the two platforms were separately imputed using the Haplotype Reference Consortium (McCarthy et al., 2016) and Minimac3 1.0.13 (Fuchsberger, Abecasis, & Hinds, 2014; Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012) available on the *Michigan Imputation Server* as reference data. A series of quality checks was performed before merging data from the two platforms' imputation (e.g. platform effects, allele frequencies by imputation quality). For the present analyses we limited our analyses to variants genotyped or imputed at info >.70 on both platforms, allele frequency difference between platforms smaller than 5%, and Hardy Weinberg *p*-value was greater than 10^{-5}. Using these criteria, 7,581,516 genotyped and well-imputed SNPs were retained for the analyses.

We performed principal component analysis on a subset of 42,859 common (MAF>5%) autosomal HapMap3 SNPs (Haplotype Reference Consortium, 2016), after stringent pruning to remove markers in linkage disequilibrium ($r^2 > 0.1$) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected.

Of the final sample of successfully genotyped individuals, there were 4,814 people who also had information on school type and exam results at age 16 which were included in the present analysis.

References

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience, 4*(1), 7.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

Fuchsberger, C., Abecasis, G. R., & Hinds, D. A. (2014). minimac2: faster genotype imputation. *Bioinformatics, 31*(5), 782-784.

Haplotype Reference Consortium. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics, 48*(10), 1279-1283.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics, 44*(8), 955-959.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., . . . Sharp, K. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics, 48*(10), 1279-1283.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria: Foundation for Statistical Computing.

Methods S2 – Creating the school type variable

To create the school type variable for the present study, we used TEDS data in combination with data from the National Pupil Database (NPD; https://www.gov.uk/guidance/national-pupil-database-apply-for-a-data-extract).

TEDS data

When the individuals in our sample were 18, they received a questionnaire that included a series of questions asking what type of school they attended during their GCSEs. Respondents were asked to indicate either ‘Yes’ or ‘No’ for different school types, including: home-school, comprehensive school, grammar school, independent (private) school, special school, sixth-form or further education college, faith school, academy and single-sex school. Respondents could select ‘Yes’ to more than one school type.

We classed all respondents who said they went to either a comprehensive or an academy school as ‘State non-selective’. Because individuals were able to select more than one school type, we excluded those who also said they went to a grammar school (n = 22), independent school (n = 26) or special school (n = 17). We did not include ‘sixth-form’ or ‘further education college’ within the state non-selective school type as we did not have any information about their selection criteria. After exclusions, the total number of individuals attending a state non-selective school was 4,780.

To create the ‘Grammar’ group, we classed all respondents who said they attended a grammar school as ‘Grammar’. Again, we excluded those who indicated that they also went to a private school (n = 24), comprehensive school (n = 22) or special school (n = 3). After exclusions, the total number of individuals in this group was 372. We classed all respondents who said they attended a private school as ‘Private’. We excluded those who indicated that they also went to a comprehensive school (n = 26), grammar school (n = 24) or special school (n = 8). After exclusions, the total number of individuals in this group was 513. We could not class individuals who indicated that they went to a faith or single sex school only into one of the three school types, as these schools can be state non-selective, grammar or private schools.

National Pupil Database data

In order to increase sample sizes, we also accessed school type information through the National Pupil Database (NPD). NPD is a pupil-level database which matches pupil and school characteristic data to pupil level attainment in England. Within the TEDS sample, 13,392 individuals gave consent for us to access their NPD records, of which 12,717 individuals were successfully matched. Approximately 700 individuals who had given consent lived outside of the England (for example Wales or Scotland), and therefore individuals could not be matched. In addition to pupil-level data on attainment, NPD also includes information on what type of school an individual attended during their GCSEs which is limited in description to one school type (for a list of school types in NPD and corresponding sample sizes in our data, please see Table SM1). Students coded in NPD as attending: 'community', 'voluntary aided', 'voluntary controlled', 'foundation', 'city technology college', 'non-maintained', 'academy sponsor-led', 'academy-converter' or 'free schools' were classed as 'State non-selective' (*n* = 10,446). Because NPD does not include a separate category for grammar schools, we identified grammar schools using the Department for Education database 'EduBase' which we could link to NPD data through unique school reference numbers (URNs). This identified 314 students attending grammar schools within our NPD records. Therefore, after excluding these individuals, there were 10,132 individuals attending 'State non-selective' schools in NPD and 314 individuals attending grammar schools. Students coded as attending 'other independent' schools in NPD we classed as 'Private' (*n* = 998).

Table SM1 – Type of establishment code taken from EduBase and sample sizes in full and selected samples

	Full sample		Selected sample	
Value label	Frequency	Percent	Frequency	Percent
Community	4031	35.0	1630	35.8
Voluntary aided	1459	12.7	551	12.1
Voluntary controlled	355	3.1	156	3.4
Foundation	2469	21.5	983	21.6
City Technology College	9	0.1	1	0.0
Community special	20	0.2	0	0.0
Other independent special	10	0.1	0	0.0
Other independent	998	8.7	386	8.5
Foundation special	2	0.0	0	0.0
Pupil referral unit	24	0.2	0	0.0
Further Education	4	0.0	0.0	0.0
Academy, sponsor-led	462	4.0	177	3.9
Academy, converter	1660	14.4	664	14.6
Free School	1	0.0	1	0.0
Total	11504	100.0	4549	100.0

Note: *Selected sample = sample who have educational achievement at 16 (GCSE), genotype data and school type data*

TEDS and NPD accuracy

There were 4186 individuals who had both TEDS data and NPD data. From this, we checked the accuracy of our groupings using descriptive crosstabs (see Table SM2). This shows the agreement between TEDS and NPD school type data. It revealed high accuracy for both the state non-selective and the private school groups. There were 75 individuals who had stated that they attended a grammar school in the TEDS data, but who actually attended a state non-selective school, as indicated by NPD. This is likely due to grammar schools converting to state non-selective schools, but keeping the title ‘grammar’ within their school name. We decided prioritise the NPD data in these cases.

Table SM2 – Cross tabulation comparison of school type data for TEDS and National Pupil Database

		TEDS: Non-selective	TEDS: Grammar	TEDS: Private	NPD accuracy
NPD	Non-selective	3473	75	8	97.67%
NPD	Grammar	2	231	1	98.71%
NPD	Private	3	3	390	98.48%
TEDS accuracy		99.86%	74.76%	97.74%	

Note: *Squares in dark grey represent individuals who were classed in NPD and TEDS as attending the same school type. Squares in light grey represent those whose school type was different in TEDS and NPD*

School type totals

After combining TEDS and NPD school type data and prioritising NPD data with relation to grammar schools, there were a total of 12,923 individuals for whom we had school type data available. 11,434 attending non-selective state schools, 377 attending grammar schools and 1112 attending private schools. The proportion of students attending the three school types in the current study is representative of UK statistics: for example grammar school UK intake = ~4%[1], our sample = 2.9%; private school UK intake = ~7%[2], our sample = 8.6%.

Of this final number 4,814 also had GCSE data and genotype information, with 4,263 attending non-selective schools, 143 attending grammar school and 408 attending private schools. 2533 people also had data for the selection factors: family SES, prior ability and prior achievement.

State non-selective schools and local education authorities

Local education authorities (LEAs) are the local councils in England and Wales that are responsible for education within their jurisdiction. They can be non-selective (contains no grammar schools), partially selective (contains one or more grammar school) or wholly selective (over 25% of pupils in that LEA attends a grammar school). Previous research suggests that those attending non-selective schools in wholly selective areas perform worse than those in non-selective areas, so we further split our ‘State non-selective’ school type into three subcategories to test this.

Non-selective, partially selective and wholly selective local education authorities (LEAs) were identified from The Education (Grammar School Ballots) Regulations 1998[3], which includes 10 ‘wholly-selective’ LEAs and a further 26 partially selective LEAs. We matched this information to our own data through school LEA.

There were 331 students attending a non-selective school in a wholly selective area, 905 students attending a non-selective school in a partially selective area, and 3,027 students attending a non-selective school in a non-selective area. Numbers for grammar (*n* = 143) and private (*n* = 408) schools remained the same.

References:

1. Bolton, P. (2017) *Grammar school statistics*. Briefing Paper 1398. House of Commons Library. Accessed at: http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN01398#fullreport

2. *Independent Schools Council Census and Annual Report (*2017). Accessed at: https://www.isc.co.uk/media/4069/isc-census-2017-final.pdf

3. The Education (Grammar School Ballots) Regulations 1998: http://www.legislation.gov.uk/uksi/1998/2876/schedule/1/made

Methods S3 – Hierarchical linear regression to calculate adjusted means for school type

To test the effect of school type on GCSE once selection factors (SES, prior achievement and prior ability) had been controlled for, we conducted hierarchical linear regression. In the first step, we entered the selection factors, which were first standardised so that the mean of these variables was 0, and in the second step of the model we entered school type. Because school type is a nominal variable with three categories (non-selective state school, grammar school and private school) without intrinsic ordering, we created two dummy coded variables to represent the three categories. This is a common way of entering nominal variables into multiple linear regression in order to capture all of the categories. Dummy coding requires one of the categories to be the reference category, in which the other categories are compared with; in this analysis we chose to use state non-selective schools as the reference category to look at the effects of selective schools on GCSE performance (see Supplementary Methods S3 for further information).

Conducting hierarchical linear regression enables us to observe the R^2 change between the two steps in the model, indicating the amount of variance in mean GCSE score explained by school type once selection factors have been controlled for. In addition, it also allows us to test whether mean GCSE score differs between school types whilst keeping the selection factors constant. For example, in the case of grammar schools, the mean would be calculated using the equation below:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Where $\hat{Y}$ is the mean GCSE for grammar schools, β_0 is the intercept in the second step of the model which, in this case, is the expected mean GCSE of state non-selective schools when all other independent variables are 0 (which have been standardised so that 0 represents their mean), X_1, X_2, X_3 and X_4 are the independent variables: school type, SES, prior ability and prior achievement and β_1, β_2, β_3 and β_4 are the beta coefficients associated with the change in dependent variable when school type goes from state non-selective school to grammar school, whilst keeping the other independent variables constant. We observed the t statistic and its associated significance in order to see whether the mean GCSE differed between groups, once accounting for selection factors.

Table S1 – Analysis of variance (ANOVA) and planned contrasts for *EduYears* GPS between students of three school types: state non-selective, grammar and private schools

				ANOVA		Contrasts			
	n	Mean (SD)	95% CIs	F	η^2	Contrasts	t	d_{cohen}	95% CI
Non-selective (N)	4263	-.043 (1.000)	-.073 to -.014			G vs N	4.869***	.413	.246-.579
Grammar (G)	143	.368 (0.989)	.204 to .531	35.800***	.014	G vs P	.436	N/A	N/A
Private (P)	408	.325 (0.954)	.233 to .418			P vs N	7.170***	.372	.270-.473

Note: n = number of participants in each group; SD = standard deviation; 95% CIs = 95% confidence intervals around the mean; F = test of overall ANOVA model; η^2 = eta squared variance explained; N = non-selective state school students; G = grammar school students; P = private school students; d_{cohen} = adjusted cohen's d statistic; CI = confidence intervals; *** = $p < .001$

Table S2 – Analysis of variance (ANOVA) and planned contrasts for *EduYears* GPS between students of five school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private schools

				ANOVA		Contrasts			
	n	Mean (*SD*)	95% CIs	*F*	η^2	Comparison	*t*	d_{cohen}	95% *CI*
State non-selective schools						WS vs PS	1.362	*N/A*	*N/A*
Wholly selective area (WS)	331	.002 (1.010)	-.107 to .112			WS vs NS	0.673	*N/A*	*N/A*
Partially selective area (PS)	905	-.084 (1.001)	-.150 to -.019	18.503***	.015	WS vs G	-3.675***	.369	(.248-.489)
Not selective area (NS)	3027	-.036 (.994)	-.072 to -.001			WS vs P	-4.398***	.330	(.184-.476)
Selective schools						PS vs NS	-1.280	*N/A*	*N/A*
Grammar (G)	143	.368 (.989)	.204 to .531			PS vs G	-5.058***	.452	(.275-.630)
Private (P)	408	.325 (.954)	.233 to .418			PS vs P	-6.923***	.415	(.297-.532)
						NS vs G	-4.752***	.407	(.239-.575)
						NS vs P	-6.908***	.365	(.261-.469)

Note: *n* = number of participants in each group; *SD* = standard deviation; 95% CIs = 95% confidence intervals around the mean; *F* = test of overall ANOVA model; η^2 = eta squared variance explained; WS = non-selective school in wholly selective area; PS = non-selective school in partially selective area; NS = non-selective school in non-selective area; N = non-selective state school students; G = grammar school students; P = private school students; d_{cohen} = adjusted Cohen's d statistic; *CI* = confidence intervals.* = *p* <.05; ** = *p* < .01; *** = *p* < .001.

Table S3 – Correlation matrix

	EduYears GPS	GCSE	Prior ability	Prior achievement	SES	School type
EduYears GPS	1					
GCSE	.277***	1				
Prior ability	.146***	.524***	1			
Prior achievement	.229***	.805***	.512***	1		
SES	.256***	.493***	.318***	.380***	1	
School type	.121***[a]	.300***[a]	.175***[a]	213***[a]	0.306***[a]	1

***Note**:* GPS = genome-wide polygenic score; GCSE = General Certificate of Secondary Education; prior ability = general cognitive ability based on verbal and non-verbal tests at age 11; prior achievement = achievement scores in English and maths at age 11; SES = socio-economic status. [a] = School type correlations obtained through regression using dummy-coded variables; *** = p<.001.

Table S4 – Hierarchical regression analysis of EduYears GPS, controlling for selection factors for students of three school types: state non-selective, grammar and private schools

	Step 1				Step 2			
	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t
Constant	.021 (-.015-.058)	.019		1.141	.007 (-.032-.045)	.020		.346
Selection factors								
SES	.172 (.132-.213)	.020	.176	8.431***	.165 (.124-.206)	.0210	.168	7.941***
Prior ability	.011 (-.032-.054)	.022	.011	.500	.009 (-.034-.052)	.022	.009	.400
Prior achievement	.152 (.107-.196)	.023	.155	6.696***	.146 (.101-.191)	.023	.149	6.375***
School types								
Non-selective vs Grammar					.202 (-.012-.415)	.109	.036	1.853
Non-selective vs Private					.145 (-0.18-.308)	.083	.034	1.739
	Model statistics - Step 1				**Model statistics - Step 2**			
R^2 (Std. Error)	.079 (.078)				.081			
R^2 change	.079				.002			
F Change	72.294***				3.007			

Note: SES = Socioeconomic status; CIs = confidence intervals; * p < 0.05; ** p < 0.01; *** p < 0.001. School type was dummy-coded into two variables with state non-selective schools as the reference category. Constant = mean of state non-selective schools when all other variables held constant; Model step 1: selection factors (SES, prior ability and prior achievement) were entered into the model; Model step 2: selection factors and school type were entered into the model together.

Table S5 – Hierarchical regression analysis of EduYears GPS, controlling for selection factors for students of five school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private schools

	Step 1				Step 2			
	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t
Constant	.021 (-.015-.058)	.019		1.141	.182 (.043-.321)	.071		2.563*
Selection factors								
SES	.172 (.132-.213)	.020	.176	8.431**	.168 (.127-.208)	.021	.171	8.069***
Prior ability	.011 (-.032-.213)	.022	.011	0.5	.007 (-.036-.050)	.022	.007	.310
Prior achievement	.152 (.107-.196)	.023	.155	6.696***	.148 (.103-.192)	.023	.151	6.451***
School types								
NS_WS vs NS_NS					-.182 (-.328--.036)	.075	-.089	-2.442*
NS_WS vs NS_PS					-.212 (-.378--.051)	.082	-.087	-2.579*
NS_WS vs G					.025 (-.228-.277)	.129	.004	.192
NS_WS vs P					-.032 (-.244-.179)	.108	-.008	-.300
	Model statistics for step 1				**Model statistics for step 2**			
R^2 (Std. Error)	.079 (0.941)				0.084 (0.939)			
R^2 change	.079				.005			
F Change	72.294***				3.252*			

Note: SES = Socioeconomic status; CIs = confidence intervals; NS_WS = State non-selective school in wholly selective area; NS_PS = State non-selective school in partially selective area; NS_NS = State non-selective school in non-selective area; G = Grammar school; P = Private school; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. School type was dummy-coded into two variables with state non-selective schools as the reference category. Constant = mean of state non-selective schools when all other variables held constant. Model step 1: selection factors (SES, prior ability and prior achievement) were entered into the model; Model step 2: selection factors and school type were entered into the model together.

Table S6 – Regression analysis of predictors of mean GCSE for three school types: state non-selective, grammar and private schools

	School type on GCSE (unadjusted)				School type on GCSE controlling for *EduYears*				School type on GCSE controlling for SES			
	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t
Constant	8.889 (8.842-8.936)	0.024		370.98***	8.893 (8.848-8.939)	0.023		383.832***	8.927 (8.885-8.969)	0.021		415.360***
School types												
N vs G	1.226 (0.969-1.483)	0.131	0.180	9.361***	1.086 (0.837-1.335)	0.127	0.159	8.547***	0.894 (0.663-1.125)	0.118	0.131	7.581***
N vs P	1.066 (0.871-1.260	0.099	0.206	10.741***	0.951 (0.762-1.140)	0.096	0.184	9.867***	0.574 (0.396-0.752)	0.091	0.111	6.313***
EduYears GPS					0.303 (0.258-0.347)	0.023	0.248	13.311***				
Selection factors												
SES									0.530 (0.488-0.571)	0.021	0.443	25.125***
Prior ability												
Prior achievement												
Model statistics												
R^2 (Std. Error)	0.071 (1.152)				0.132 (0.113)				0.257 (1.030)			
R^2 change	0.071				0.056				0.027			
F Change	97.243***				81.099***				45.665***			

Note: *Table continues on the next page*

	School type on GCSE controlling for ability				School type on GCSE controlling for achievement				School type on GCSE controlling for everything			
	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t
Constant	8.919 (8.879-8.960)	0.021		432.032***	8.947 (8.918-8.976)	0.015		604.673***	8.960 (8.933-8.987)	0.014		646.634***
School types												
N vs G	0.875 (0.653-1.097)	0.113	0.128	7.721***	0.259 (0.98-0.420)	0.082	0.038	3.159**	0.180 (0.30-0.331)	0.077	0.026	2.346*
N vs P	0.728 (0.559-0.897)	0.086	0.141	8.445***	0.571 (0.450-0.692)	0.062	0.110	9.266***	0.362 (0.246-0.477)	0.059	0.070	6.155***
***EduYears* GPS**									0.079 (0.052-0.107)	0.014	0.065	5.649***
Selection factors												
SES									0.194 (0.164-0.223)	0.015	0.162	13.038***
Prior ability	0.595 (0.556-0.634)	0.020	0.498	29.787***					0.150 (0.120-0.181)	0.016	0.126	47.162***
Prior achievement					0.928 (0.900-0.957)	0.014	0.777	64.265***	0.767 (0.735-0.799)	0.016	0.642	9.639***
Model statistics												
R^2 (Std. Error)	0.312 (0.991)				0.647 (0.710)				0.692 (0.664)			
R^2 change	0.034				0.013				0.005			
F Change	61.939***				46.191***				20.726***			

Note: SES = Socioeconomic status; CIs = confidence intervals; School type was dummy-coded into two variables with state non-selective schools as the reference category. **Constant** = mean of state non-selective schools when all other variables held constant. N = non-selective state school; G = grammar school; P = private school. **Model statistics**: R^2 = total variance explained by all of the predictors in the model; R^2 change = additional variance added by school type over and above other predictors; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table S7 – Regression analysis of predictors of mean GCSE for three school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private schools

	School type on GCSE (unadjusted)				School type on GCSE controlling for EduYears				School type on GCSE controlling for SES			
	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t
Constant	8.730 (8.560-8.899)	0.087		100.887***	8.698 (8.534-8.862)	0.084		103.961***	8.874 (8.722-9.027)	0.078		114.233***
School types												
NS_WS vs NS_NS	0.178 (0.357--0.089)	0.091	0.072	1.958	0.215 (0.042-0.388)	0.088	0.086	2.439*	0.062 (-0.099-0.222)	0.082	0.025	0.755
NS_WS vs NS_PS	0.154 (0.351--0.044)	0.100	0.052	1.531	0.203 (0.012-0.393)	0.097	0.068	2.084*	0.043-0.0134-0.219)	0.090	0.014	0.476
NS_WS vs G	1.3851.689-0.170)	0.155	0.203	8.932***	1.281 (0.987-1.575)	0.150	0.188	8.534***	0.9470.673-1.222)	0.140	0.139	6.769***
NS_WS vs P	1.225 (1.479-0.198)	0.129	0.237	9.466***	1.146 (0.900-1.391)	0.125	0.221	9.148***	0.628 (0.396-0.860)	0.118	0.121	5.305***
EduYears GPS					0.305 (0.260-0.349)	0.023	0.250	13.391				
Selection factors												
SES									0.529	0.021	0.443	25.035***
Prior ability												
Prior achievement												
Model statistics												
R^2 (Std. Error)	0.073 (1.151)				0.134 (1.113)				0.257 (1.031)			
R^2 change	0.073				0.058				0.027			
F Change	49.619***				42.106***				22.973***			

Note: Table continues on the next page

	School type on GCSE controlling for prior ability				School type on GCSE controlling for achievement				School type on GCSE controlling for everything			
	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t	B (95% CIs)	Std. Error	Beta	t
Constant	8.738 (8.592-8.883)	.074		117.421***	8.926 (8.821-9.030)	.053		166.980***	8.938 (8.560-8.899)	0.050		178.166***
School types												
NS_WS vs NS_NS	0.208 (0.055-0.362)	0.078	0.084	2.660**	0.016	0.056	0.006	0.284	0.019 (-0.085-0.122)	0.053	0.007	0.353
NS_WS vs NS_PS	0.159 (-0.010-0.329)	0.086	0.053	1.843	0.048	0.062	0.016	0.778	0.040 (-0.074-0.154)	0.058	0.013	0.690
NS_WS vs G	1.056 (0.794-1.319)	0.134	0.155	7.892***	0.281	0.097	0.041	2.893**	0.202 (0.026-0.024)	0.091	0.030	2.223*
NS_WS vs P	0.909 (0.690-1.128)	0.112	0.176	8.127***	0.593	0.080	0.115	7.373***	0.384 (0.234-0.534)	.076	0.074	5.029***
EduYears GPS									0.080 (0.052-0.107)	0.014	0.065	5.671***
Selection factors												
SES									0.193 (0.164-0.222)	0.015	0.162	12.984***
Prior ability	0.596	0.020	0.499	29.856***					0.150 (0.120-0.181)	0.016	0.126	9.621***
Prior achievement					0.928	0.014		64.181***	0.767 (0.735-0.799)	0.016	0.642	47.073***
Model statistics												
R^2 (Std. Error)	0.315 (0.990)				0.647 (0.710)				0.692 (0.664)			
R^2 change	0.036				0.013				0.005			
F Change	32.889***				23.331***				10.510***			

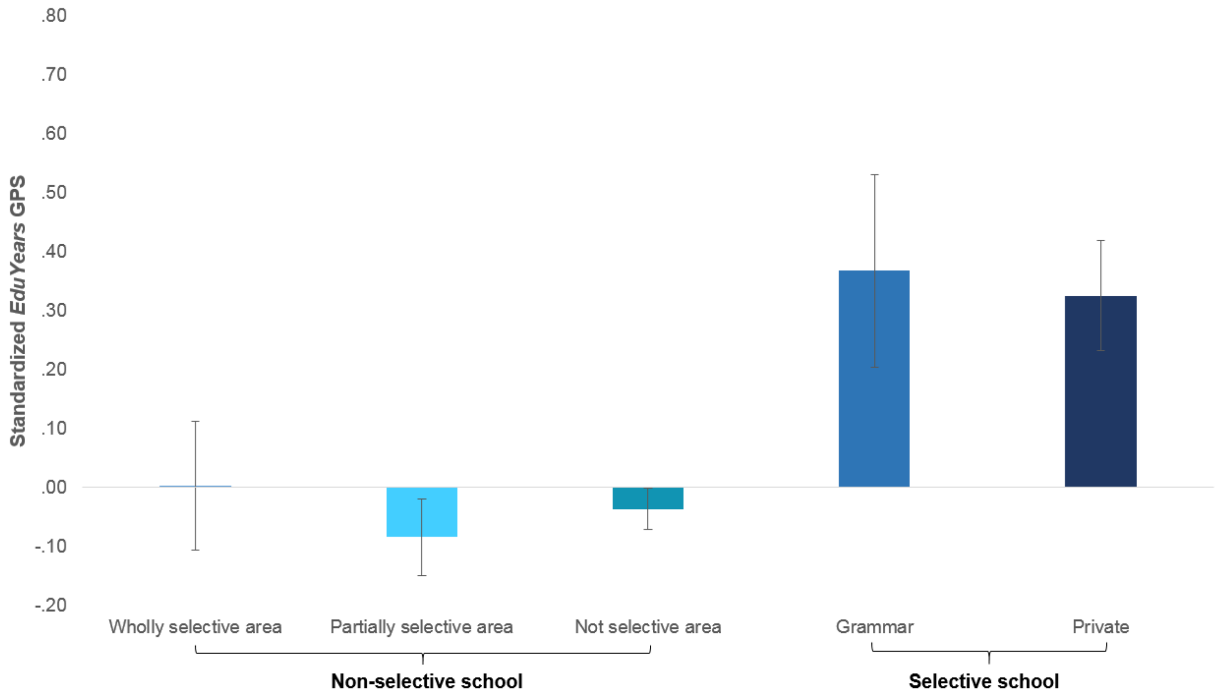
Note: SES = Socioeconomic status; CIs = confidence intervals; School type was dummy-coded into two variables with state non-selective schools as the reference category. **Constant** = mean of state non-selective schools in wholly selective area when all other variables held constant; NS_WS = State non-selective school in wholly selective area; NS_PS = State non-selective school in partially selective area; NS_NS = State non-selective school in non-selective area; G = Grammar school; P = Private school. **Model statistics**: R^2 = total variance explained by all of the predictors in the model; R^2 change = additional variance added by school type over and above other predictors; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table S8 – Descriptive statistics

	Whole sample			Non-selective schools			Grammar schools			Private schools		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
EduYears GPS	4814	.04	1.00	4263	.00	.99	143	.41	.99	408	.36	.95
GCSE	4814	8.84	1.26	4263	8.7	1.24	143	10.01	.77	408	9.85	.96
SES	4574	.20	.98	4034	.09	.95	136	.72	.91	404	1.09	.73
Prior ability[1]	2922	.06	.97	2582	.00	.98	96	.61	.69	244	.49	.78
Prior achievement	4298	68.77	15.63	3935	67.82	15.63	123	84.50	6.17	240	76.32	12.34

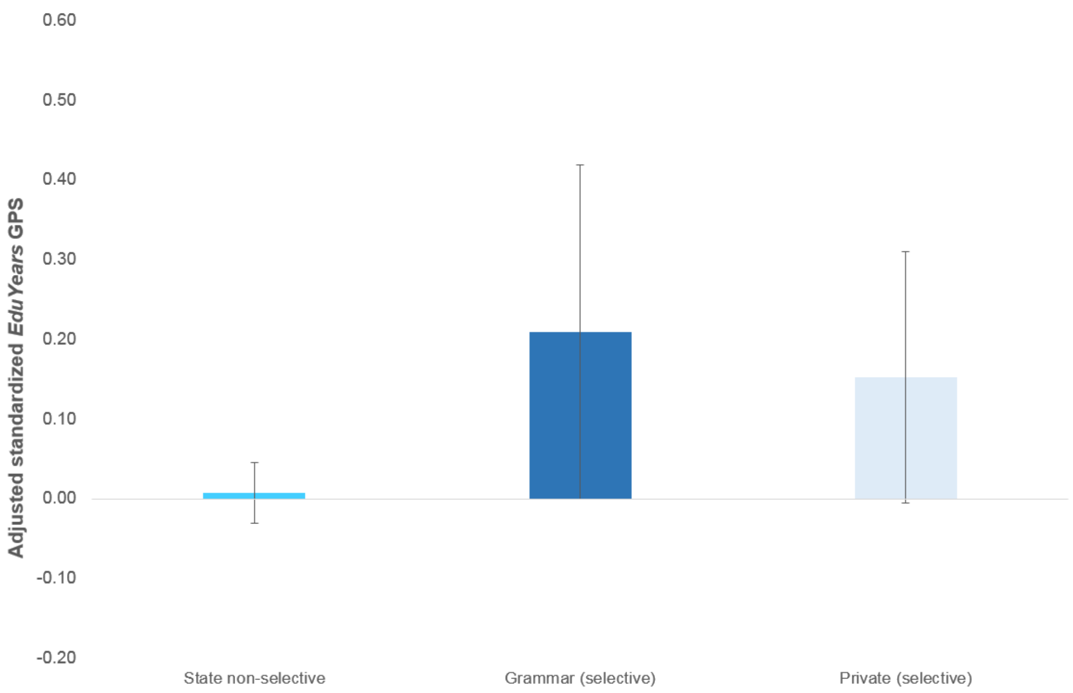
Note: *N = number of participants; SD = standard deviation; GCSE = General Certificate of Secondary Education; GPS = genome-wide polygenic score; SES = socioeconomic status. [a] = Full sample using one twin in a pair randomly to maintain independence of data; [1] = standardisation was required to form a composite. For those measures that were standardised, standardization was done on the full sample to show comparison to the selected sample and to the different school types. Descriptives were computed with raw data.*

Figure S1 – *EduYears* GPS plotted means (and 95% confidence intervals) for students of five school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private schools



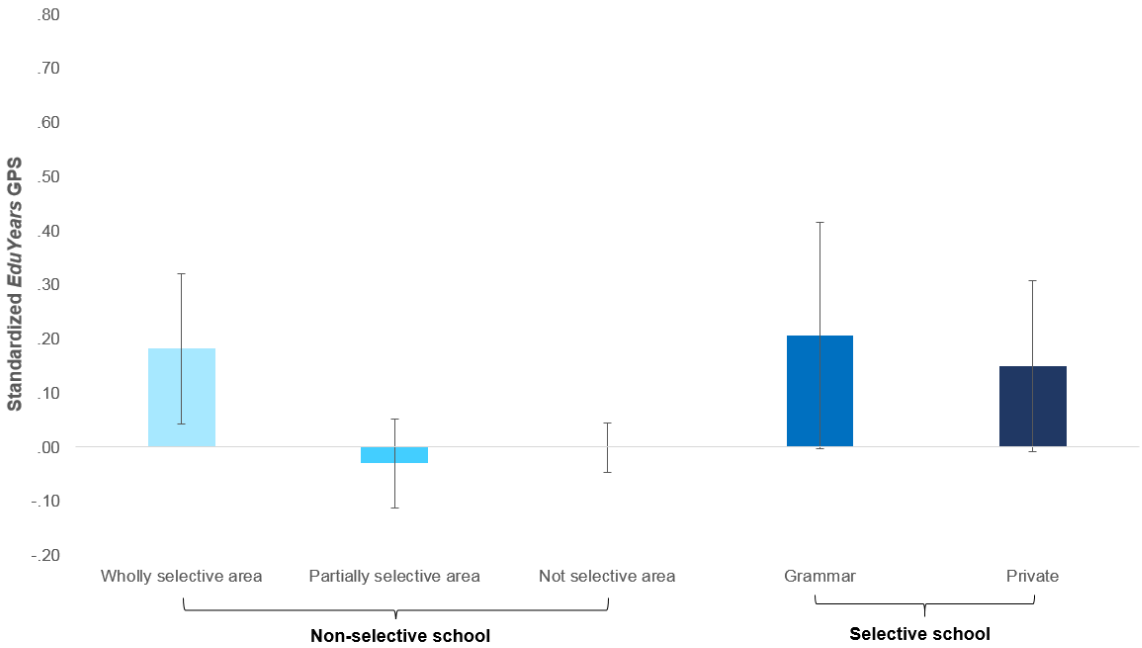
Note: There were no significant mean differences for *EduYears* GPS between state non-selective school students in varying selectively areas (wholly selective vs partially selective t= 1.362, p = .173; wholly selective vs not selective area t= 0.673, p =.501; partially selective vs not selective area t= -1.280, p = .200). There were significant mean differences between all of the state non-selective groups and both grammar and private school students (state wholly selective vs grammar t= -3.675, p <.001; state wholly selective vs private t = -4.398, p <.001; state partially selective vs grammar t= -5.058, p <.001; state partially selective vs private t = -6.923, p <.001; state non selective area vs grammar t= -4.752, p <.001; state not selective area vs private t= -6.908, p <.001).

Figure S2 – *EduYears* GPS plotted means (and 95% confidence intervals) controlling for selection factors between students of 3 school types: non-selective state, grammar and private



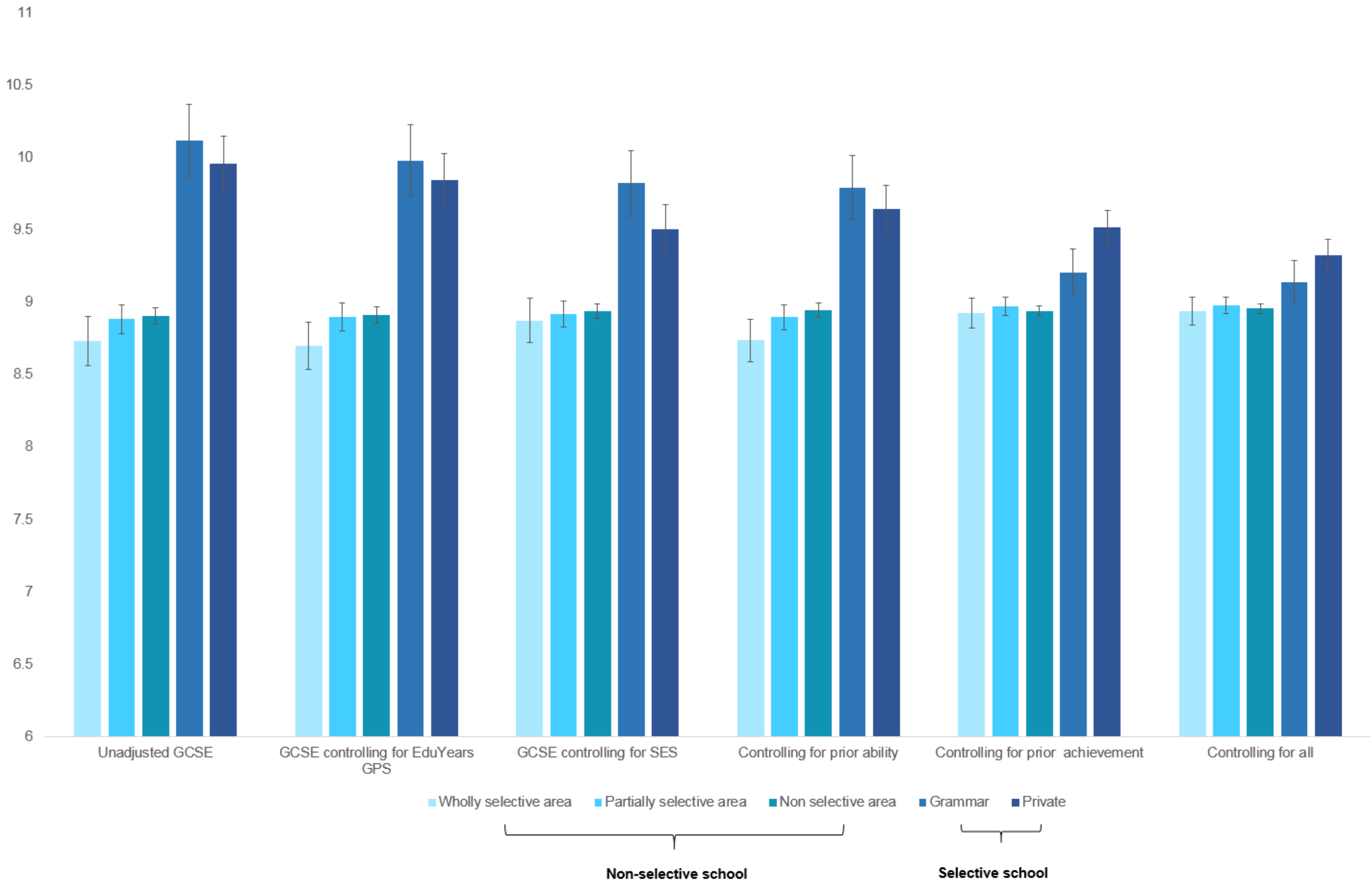
Note: There were no significant *EduYears* GPS mean differences between state non-selective and grammar school students (t = 1.853, p = 0.064) or between state non-selective and private school students (t = 1.739, p = 0.082) or between grammar and private school students (t = .432, p = 0.665). The 95% confidence intervals are larger here than in Figure 1 because the sample sizes were reduced when data for the three selection factors were required (N = 2533).

Figure S3 – *EduYears* GPS plotted means (and standard errors) controlling for selection factors between 5 school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private school



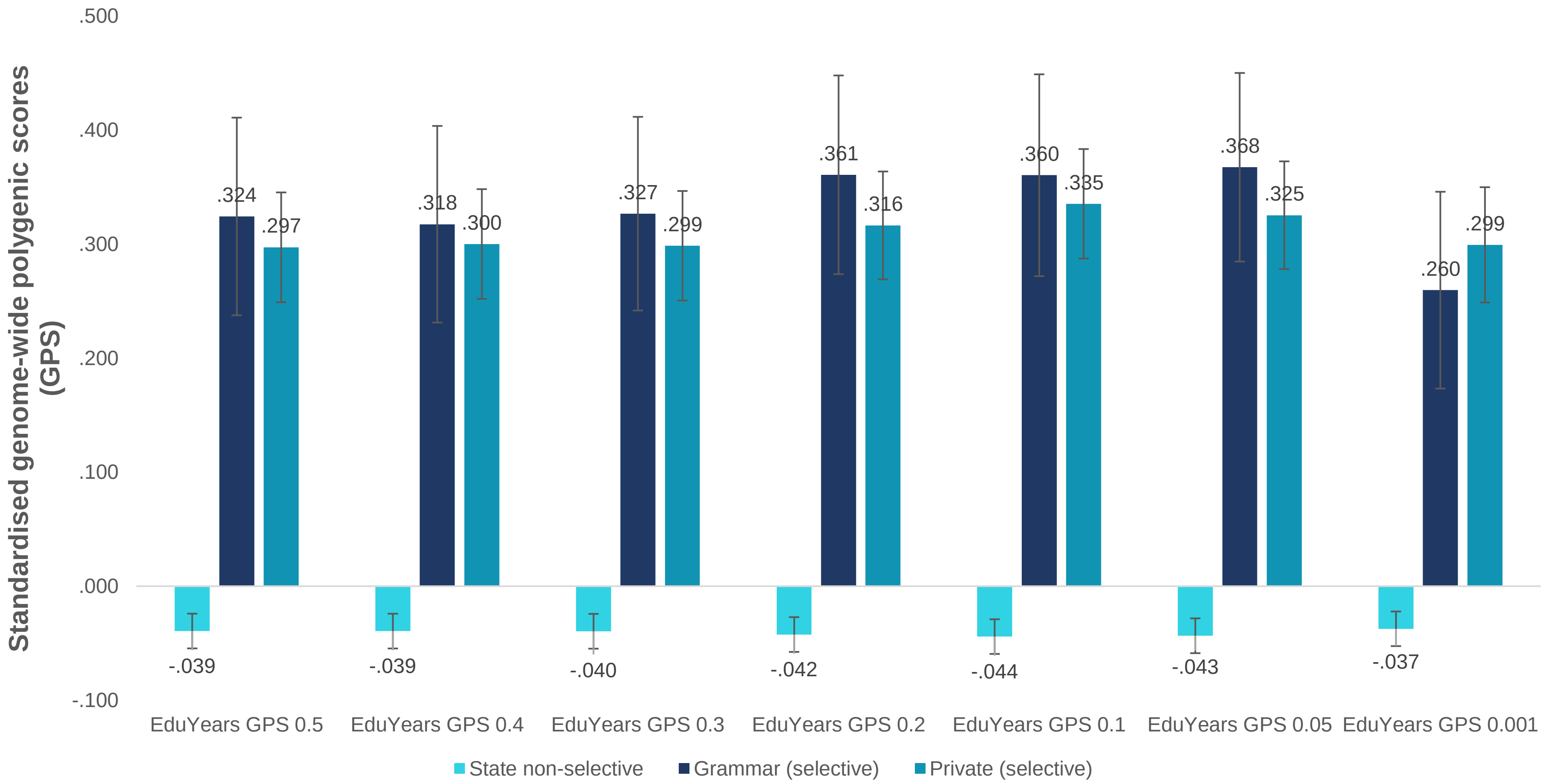
Note: There were small significant differences between students in state non-selective schools in wholly-selective vs partially selective areas (t= -2.579, p =.010) and students in wholly selective areas vs non-selective area (t= -2.442, p = .015), controlling for selection factors. The 95% confidence intervals are larger here than in Figure S1 because the sample sizes were reduced when data for the three selection factors were required (N = 2533).

Figure S4 – The plotted means (and 95% confidence intervals) for unadjusted GCSE, GCSE controlling for *EduYears* GPS, GCSE controlling for SES, GCSE controlling for prior ability, GCSE controlling for prior achievement and GCSE controlling for all variables between 5 school types: non-selective schools in wholly selective areas, non-selective schools in partially selective areas, non-selective schools in non-selective areas, grammar schools and private school



Note: For GCSE controlling for all the variables, there were no differences between non-selective school students in varying selectivity areas. However, there were differences between wholly-selective and both grammar (t = 2.223, p = .026) and private (t = 5.029, p <.001) and between partially selective areas and both grammar (t = 1.997, p = .046) and private (t = 5.348, p <.001) and non-selective and both grammar (t = 2.375, p = .018) and private (t = 6.146, p <.001).

Figure S5 – Mean *EduYears* GPS (and 95% confidence intervals) between state non-selective, grammar and private school for several *EduYears* GPS p-value cut-off



Appendix 3: Supplementary Material for Chapter 4

Ofsted secondary school quality is a poor predictor of student academic achievement and wellbeing

Smith-Woolley, E., Cheeseman, R., Pingault, J B., von Stumm, S., Asbury, K., Dale, P. S., Allen, R., Kovas, Y., & Plomin, R.

Supplementary Measures:

Measures S1 – Measures of the school engagement

Measures S2 – Measures of academic wellbeing

Supplementary Tables:

Table S1. Analysis of variance with polynomial trend analysis and planned contrasts of GCSE scores between students attending schools rated as: Outstanding, Good, Requires Improvement and Inadequate

Table S2. Results from multiple regression analysis predicting examination results at age 16 (GCSEs) from student covariates and Ofsted Headline Quality Rating.

Table S3. Analysis of covariance of GCSE scores between students attending schools rated as: Outstanding, Good, Requires Improvement and Inadequate, accounting for covariates of prior achievement and socioeconomic status

Table S4. Analysis of variance with polynomial trend analysis of school engagement and wellbeing measures between students attending schools rated as: Outstanding, Good, Requires Improvement and Inadequate

Table S5: Representativeness of the current sample

Table S6. Sample sizes, means and standard deviations (SD) for Ofsted ratings.

Table S7. Principal Component Analysis of Ofsted items

Table S8. Ofsted individual item loadings

Supplementary Figures:

Figure S1. Spearman correlation coefficients (rs) for the relationship between measures of student wellbeing and engagement and the Ofsted headline quality rating.

Figure S2. Spearman correlation coefficients for the relationship between measures of student wellbeing and engagement and the Ofsted headline quality rating

Figure S3. Scree plot

Supplementary Measures:

At age 16, participants completed 8 questionnaires about school engagement and 6 questionnaires relating to academic wellbeing. A description of each of the questionnaires is included below. Measures were collected via web tests. All measures are self-report.

Supplementary Measures S1 – Measures of the school engagement

Teacher-student relations – 6 items (Appleton, Christenson, Kim, & Reschly, 2006): This is a subscale of the Student Engagement Instrument and included items such as: "At my school, teachers care about students" and "My teachers are there for me when I need them" - rated on a 4-point scale from strongly disagree to strongly agree. The total score was created by taking the mean of the 6 items, requiring at least 3 to be present for an individual. The reported reliability of this subscale is good (alpha = .88).

Control relevance of school work – 4 items (Appleton et al., 2006): This is a subscale of the Student Engagement Instrument and included items such as: "I feel like I have a say about what happens to me at school" and "When I do well in school, it's because I work hard" - rated on a 4-point scale from strongly disagree to strongly agree. The total score was created by taking the mean of the 4 items, requiring at least 2 to be present for an individual. The reported reliability of this subscale is good (alpha = .80).

Peer support for learning – 3 items (Appleton et al., 2006): This is a subscale of the Student Engagement Instrument and included items such as: "Students at my school respect what I have to say" and "Students at my school are there for me when I need them" rated on a 4-point scale from strongly disagree to strongly agree. The total score was created by taking the mean of the 3 items, requiring at least 2 to be present for an individual. The reported reliability of this subscale is good (alpha = .82).

Family support for learning – 3 items (Appleton et al., 2006): This is a subscale of the Student Engagement Instrument and included items such as: "When something good happens at school, my family/carer(s) want to know about it." and "My family/carer(s) want me to keep trying when things are tough at school." rated on a 4-point scale from strongly disagree to strongly agree. The total score was created by taking the mean of the 3 items, requiring at least 2 to be present for an individual. The reported reliability of this subscale is good (alpha = .76).

Homework behaviour – 2 items (Programme for International Student Assessment, 2001). These questions were taken from the PISA 2000, 2003 and 2006 student questionnaires. For homework behaviour, 2 items were selected: “I complete my homework on time” and “I do my homework while watching television” (reversed). These questions were rated on a 4-point scale from ‘never’ to ‘always’. A mean of these two items was taken as the total score for an individual, requiring both items to be present.

Homework feedback – 3 items (Programme for International Student Assessment, 2001). These questions were taken from the PISA 2000, 2003 and 2006 student questionnaires. For homework feedback, 3 items were selected: “My teachers grade my homework”, “My teachers make useful comments on my homework” and “I am given interesting homework”. These questions were rated on a 4-point scale from ‘never’ to ‘always’. A mean of these three items was taken as the total score for an individual, requiring at least two items to be present.

Attitudes to school – 4 items (Programme for International Student Assessment, 2001). These questions were taken from the PISA 2000, 2003 and 2006 student questionnaires. For this measure, four questions were asked relating to attitudes to the school, such as “School has done little to prepare me for adult life when I leave school” (reversed) and “School has taught me things which could be useful in a job”. These four questions were rated on a 4-point scale from strongly disagree to strongly agree. A mean of these items was taken as the total score for an individual, requiring at least 2 items to be present.

Peer victimisation – 6 items (Mynard & Joseph, 2000). These questions were taken from the Multidimensional Peer-Victimization Scale which measures physical and verbal victimisation as well as social manipulation and attacks on property. Participants were asked to indicate whether and how often another student had victimised them, for example “How often during this school year has another student made fun of me for some reason” or “Hurt me physically in some way?” The response options were: ‘not at all’, ‘once’ or ‘more than once’. A mean of the items was used as a total score requiring at least half of the items to be present.

Supplementary Measures S1 – Academic wellbeing

Academic self-concept – 10 items (Burden, 1998): These questions were taken from the 'Myself-As-Learner Scale' which was developed to measure academic self-concept in secondary-school aged learners. Participants were required to indicate the extent to which a series of statements describe them. These statements included things like "When I get stuck with my work I can usually work out what to do next" and "When I'm given new work to do, I usually feel confident I can do it". There was a 5-point rating scale from 'Very much like me' to 'Not like me at all'. A mean of the 10 items was taken as a total score requiring at least half to be present for each individual.

Future aspirations and goals – 3 items (Appleton et al., 2006): This is a subscale of the Student Engagement Instrument and included the following items: "I plan to continue my education following school", "School is important for achieving my future goals" and "I am hopeful about my future." rated on a 4-point scale from strongly disagree to strongly agree. The total score was created by taking the mean of the 6 items, requiring at least 3 to be present for an individual. The reported reliability of this subscale is good (alpha = .78).

Life satisfaction in relation to school – 4 items (Huebner, 1994). This is a subscale of the Multidimensional Students' Life Satisfaction Scale. It included items tapping into life satisfaction, with a focus on the school environment such as "I like being in school" and "I enjoy school activities". Participants were asked to indicate to what extent they agreed with these statements using a 6 point scale from 'Strongly agree' to 'Strongly disagree'. The total score was created by taking a mean of the items, requiring at least half to be present.

Subjective happiness – 4 items (Lyubomirsky & Lepper, 1999).This measure requires students to rate themselves on a 7 point scale for statements such as "In general, I consider myself to be"… (1) 'not a very happy person' to (7) 'a very happy person' or "Some people are generally very happy. They enjoy life regardless of what is going on, getting the most out of everything. To what extent does this describe you?" From (1) 'not at all' to (7) 'a great deal'. The total score was created by taking a mean of the four items, requiring at least half to be present.

Grit – 9 items (Duckworth & Quinn, 2009): This measure required participants to rate statements such as 'I am driven to succeed' on a 5-point scale from 'very much like me' to ‘not like me at all’. The total score was created by taking the mean of the 9 items, requiring at least 5 to be present.

Ambition – 5 items (Duckworth & Quinn, 2009): This measure required participants to rate statements such as ‘I aim to be the best in the world at what I do’ and ‘I am ambitious’ on a 5-point scale from ‘very much like me’ to ‘not like me at all’. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present.

References:

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of school psychology, 44*(5), 427-445.

Burden, R. (1998). Assessing children's perceptions of themselves as learners and problem-solvers: The construction of the Myself-as-Learner Scale (MALS). *School Psychology International, 19*(4), 291-305.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment, 91*(2), 166-174.

Huebner, E. S. (1994). Preliminary development and validation of a multidimensional life satisfaction scale for children. *Psychological Assessment, 6*(2), 149-158.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research, 46*(2), 137-155.

Mynard, H., & Joseph, S. (2000). Development of the multidimensional peer-victimization scale. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression, 26*(2), 169-178.

Programme for International Student Assessment. (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*: Organisation for Economic Co-operation and Development.

Supplementary Tables

Table S1. Analysis of variance with polynomial trend analysis and planned contrasts of GCSE scores between students attending schools rated as: Outstanding, Good, Requires Improvement and Inadequate

ANOVA					
	SS	*df*	*F*	*p*	η^2
Between groups					
Combined	296.37	3	67.47	1.22 x 10-42	
Linear	295.73	1	201.96	7.68 x 10-45	
Quadratic	0.55	1	0.38	0.54	
Cubic	0.09	1	0.06	0.81	
Within groups	6394.53	4367			
Total	6690.9	4370			0.44

Planned contrasts				
	Inadequate	Requires Improvement	Good	Outstanding
Inadequate *M*=8.17 (*SD* = 1.24)		3.10*	6.49**	9.93**
Requires Improvement *M*=8.47 (*SD* = 1.23)	0.30		6.35**	12.31**
Good *M*=8.77 (*SD* = 1.21)	0.60	0.30		7.78**
Outstanding *M*=9.11 (*SD* = 1.24)	0.94	0.64	0.34	

Note: *SS* = Sum of squares; *df* = degrees of freedom; *F* = test of overall ANOVA model; *p* = significance of F statistic; η^2 = eta squared variance explained; Planned contrasts: *M* = mean; *SD* = standard deviation; the lower diagonal matrix shows the mean GCSE grade differences between Ofsted categories. The upper diagonal matrix presents the *t* static associated with the difference. ** = *p* <.001 * = *p* <.05.

Table S2. Results from multiple regression analysis predicting examination results at age 16 (GCSEs) from student covariates and Ofsted Headline Quality Rating.

Predictors	*B*	*SE*	*β*	*t*	*p*	*r*	*sr*
Socioeconomic status	.197 (.170-.224)	.014	.156	14.298	1.87×10^{-44}	.444	.143
KS2 English score	.331 (.294-.368)	.019	.262	17.236	5.25×10^{-64}	.713	.173
KS2 Mathematics score	.462 (.422-.502)	.020	.376	22.647	9.41×10^{105}	.759	.228
KS2 Science score	.232 (.189-.275)	.022	.186	10.563	7.98×10^{-26}	.730	.106
Headline Quality Rating	.127 (.097-.157)	.015	.085	8.234	$2.80 \times 10\text{-}^{16}$	.211	.083

Full model statistics:

$F(5, 3007) = 1379.093$
$p = <.000001$
$R^2 = .696$

Note: Beta coefficients, standard errors and t statistics, p-values and correlations are presented for each of the predictors in the multiple regression model. *β* = standardised Beta coefficient; *SE* = standard error; t = unstandardised beta coefficient divided by the *SE*, *p* = significance of result; *r* = Pearson correlation between predictor and GCSE; *sr* = semi-partial correlation - unique prediction of predictor on GCSE corrected for other predictors in the model; *F* = test of overall ANOVA model; R^2= variance explained by all the predictors in the model.

Table S3. Analysis of covariance of GCSE scores between students attending schools rated as: Outstanding, Good, Requires Improvement and Inadequate, accounting for covariates of prior achievement and socioeconomic status

ANOVA					
	SS	*df*	*F*	*p*	η^2
KS2 English	131.22	1	297.29	1.35×10^{-63}	0.03
KS2 maths	226.39	1	512.93	5.79×10^{-105}	0.05
KS2 Science	49.22	1	111.51	1.28×10^{-25}	0.01
Socioeconomic status	88.56	1	200.65	3.87×10^{-44}	0.02
Ofsted-rating	30.22	3	22.82	1.33×10^{-14}	<.01
Error	1323.21	2998			
Total	4355.16	3005			

Pairwise comparisons				
	Inadequate	Requires Improvement	Good	Outstanding
Inadequate *M*=8.55 (*SE* = 0.06)		0.07	1.60×10^{-5}	2.91×10^{-9}
Requires Improvement *M*=8.72 (*SE* = 0.02)	0.17		1.97×10^{-4}	6.40×10^{-11}
Good *M*=8.85 (*SE* = 0.02)	0.30	0.13		.001
Outstanding *M*=8.96 (*SE* = 0.02)	0.41	0.24	0.11	

Note: *SS* = Sum of squares; *df* = degrees of freedom; *F* = test of overall model; *p* = significance of F statistic; η^2 = eta squared variance explained; Pairwise comparisons: *M* = mean; *SE* = standard error; the lower diagonal matrix shows the mean GCSE grade differences between Ofsted categories, once accounting for student covariates. The upper diagonal matrix presents the significance of the results. Only the difference between Inadequate and Requires Improvement was not significant.

Table S4. Analysis of variance with polynomial trend analysis of school engagement and wellbeing measures between students attending schools rated as: Outstanding, Good, Requires Improvement and Inadequate

		SS	*df*	*F*	*p*	η^2
Teacher-Student Relations	*Between Groups*	3.92	3	2.42	0.06	
	Linear	3.90	1	7.21	0.01	
	Quadratic	0.01	1	0.01	0.91	
	Cubic	0.01	1	0.02	0.89	
	Within Groups	974.95	1802			
	Total	978.87	1805			0.00
Control/Relevance of School Work	*Between Groups*	2.12	3	1.55	0.20	
	Linear	1.61	1	3.54	0.06	
	Quadratic	0.35	1	0.76	0.38	
	Cubic	0.16	1	0.36	0.55	
	Within Groups	819.61	1802			
	Total	821.73	1805			0.00
Peer Support for Learning	*Between Groups*	3.88	3	2.25	0.08	
	Linear	3.16	1	5.50	0.02	
	Quadratic	0.68	1	1.18	0.28	
	Cubic	0.05	1	0.08	0.78	
	Within Groups	1031.82	1797			
	Total	1035.70	1800			0.00
Homework Behaviour scale	*Between Groups*	18.08	3	5.10	0.00	
	Linear	8.91	1	7.53	0.01	
	Quadratic	5.13	1	4.34	0.04	
	Cubic	4.04	1	3.41	0.06	
	Within Groups	2139.09	1809			
	Total	2157.16	1812			0.01
Homework Feedback scale	*Between Groups*	20.91	3	2.23	0.08	
	Linear	19.21	1	6.14	0.01	
	Quadratic	1.70	1	0.55	0.46	
	Cubic	0.00	1	0.00	1.00	
	Within Groups	5633.07	1801			
	Total	5653.98	1804			0.00
Attitudes to School	*Between Groups*	2.23	3	2.25	0.08	
	Linear	1.91	1	5.81	0.02	
	Quadratic	0.17	1	0.52	0.47	
	Cubic	0.14	1	0.43	0.51	
	Within Groups	596.70	1810			
	Total	598.92	1813			0.00
Family Support for Learning	*Between Groups*	3.93	3	1.42	0.24	
	Linear	1.41	1	1.53	0.22	
	Quadratic	0.01	1	0.01	0.91	
	Cubic	2.51	1	2.71	0.10	
	Within Groups	1659.09	1797			
	Total	1663.02	1800			0.00

Peer Victimisation	*Between Groups*	27.18	3	0.86	0.46	
	Linear	5.31	1	0.50	0.48	
	Quadratic	6.19	1	0.59	0.44	
	Cubic	15.68	1	1.49	0.22	
	Within Groups	18924.74	1793			
	Total	18951.92	1796			0.00
Academic Self-Concept	*Between Groups*	0.89	3	0.79	0.50	
	Linear	0.44	1	1.18	0.28	
	Quadratic	0.30	1	0.79	0.37	
	Cubic	0.15	1	0.40	0.53	
	Within Groups	635.54	1692			
	Total	636.43	1695			0.00
Future Aspirations/Goals	*Between Groups*	3.19	3	1.13	0.33	
	Linear	2.72	1	2.90	0.09	
	Quadratic	0.04	1	0.04	0.84	
	Cubic	0.44	1	0.46	0.50	
	Within Groups	1683.47	1798			
	Total	1686.66	1801			0.00
Life Satisfaction School	*Between Groups*	6.42	3	2.18	0.09	
	Linear	5.91	1	6.02	0.01	
	Quadratic	0.02	1	0.02	0.89	
	Cubic	0.49	1	0.50	0.48	
	Within Groups	1850.94	1884			
	Total	1857.36	1887			0.00
Subjective happiness	*Between Groups*	2.65	3	0.64	0.59	
	Linear	2.39	1	1.73	0.19	
	Quadratic	0.25	1	0.18	0.67	
	Cubic	0.01	1	0.01	0.92	
	Within Groups	2593.72	1886			
	Total	2596.36	1889			0.00
GRIT	*Between Groups*	0.18	3	0.18	0.91	
	Linear	0.00	1	0.00	0.98	
	Quadratic	0.05	1	0.15	0.70	
	Cubic	0.13	1	0.39	0.53	
	Within Groups	577.00	1712			
	Total	577.18	1715			0.00
Ambition	*Between Groups*	2.01	3	1.48	0.22	
	Linear	1.65	1	3.65	0.06	
	Quadratic	0.06	1	0.12	0.73	
	Cubic	0.30	1	0.67	0.41	
	Within Groups	765.86	1690			
	Total	767.87	1693			0.00

Note: *SS* = Sum of squares; *df* = degrees of freedom; *F* = test of overall ANOVA model; *p* = significance of F statistic; η^2 = eta squared variance explained.

Table S5: Representativeness of the current sample

Achievement	UK population	Current sample
5 + GCSEs A* - C grade	75%	81%
Socioeconomic variables		
Mother employed	49%	48%
Father employed	89%	93%

Note: We used the 2001 UK Census data for socioeconomic variables as this was taken at the time our variables were collected: https://www.ons.gov.uk/census/2001censusandearlier/aboutcensus2001; We used the 2011 GCSE statistics as this within the period the current sample took their GCSEs: https://www.gov.uk/government/statistics/gcse-and-equivalent-attainment-by-pupil-characteristics-in-england-2010-to-2011

Table S6. Sample sizes, means and standard deviations (SD) for Ofsted ratings

	Total *N*	Mean	*SD*
Overall headline school quality measure:			
Overall effectiveness: how good is the school?	4391	2.97	0.82
Individual ratings:			
The effectiveness of partnerships in promoting learning and well-being	2903	3.29	.657
The schools capacity for sustained improvement	1722	2.99	.742
Outcomes for individuals and groups of pupils/children	3282	2.79	.797
Pupils achievement and the extent to which they enjoy their learning	2903	3.00	.897
Pupils attainment	2897	3.09	.669
The quality of pupils learning and their progress	2895	3.12	.683
The quality of learning for pupils with special educational needs and/or disabilities and their progress	2903	3.37	.636
The extent of pupils spiritual, moral, social and cultural development	2903	3.32	.610
The extent to which pupils adopt healthy lifestyles	2903	3.51	.563
The extent to which pupils feel safe	2897	3.09	.786
Pupils attendance	4385	3.15	.693
Pupils behaviour	2903	3.52	.604
The extent to which pupils contribute to the school and wider community	2903	3.27	.707
The extent to which pupils develop workplace and other skills that will contribute to their future economic well-being	4391	2.86	.702
The quality of teaching	2903	3.34	.612
The extent to which the curriculum meets pupils needs, including, where relevant, through partnerships	2903	3.56	.581
The effectiveness of care, guidance and support	1741	2.64	.610
The use of assessment to support learning	4391	3.10	.756
The effectiveness of leadership and management in embedding ambition and driving improvement	2897	3.30	.675
The effectiveness with which the school promotes equality of opportunity and tackles discrimination	2389	3.06	.695
The effectiveness with which the school promotes community cohesion	2897	3.20	.737
The effectiveness with which the school deploys resources to achieve value for money	2897	3.19	.672

The effectiveness of the governing body in challenging and supporting the school so that weaknesses are tackled decisively	1794	3.15	.600
The effectiveness of safeguarding procedures	1722	3.04	.604
The effectiveness of the schools engagement with parents and carers	1722	2.98	.661
The leadership and management of teaching and learning	1722	2.98	0.66

Table S7. Principal Component Analysis of Ofsted items

	Eigenvalues		
Component	**Total**	**% of Variance**	**Cumulative %**
1	16.054	59.461	59.461
2	1.378	5.103	64.563
3	1.206	4.468	69.032
4	.830	3.074	72.106
5	.697	2.583	74.689
6	.630	2.332	77.021
7	.606	2.246	79.267
8	.581	2.153	81.420
9	.464	1.718	83.138
10	.449	1.664	84.801
11	.442	1.636	86.438
12	.420	1.557	87.994
13	.372	1.378	89.372
14	.368	1.363	90.735
15	.349	1.292	92.027
16	.329	1.220	93.246
17	.304	1.125	94.372
18	.267	.987	95.359
19	.244	.904	96.263
20	.236	.873	97.136
21	.216	.800	97.937
22	.167	.618	98.555
23	.147	.546	99.101
24	.086	.318	99.419
25	.076	.280	99.699
26	.056	.209	99.908
27	.025	.092	100.000

Table S8. Ofsted individual item loadings

Ofsted items	Component 1	2	3
The effectiveness of partnerships in promoting learning and well-being	.76		
The school's capacity for sustained improvement	.84		
Outcomes for individuals and groups of pupils/children	.92		
Pupils achievement and the extent to which they enjoy their learning	.89		
Pupils attainment	.72	-.49	
The quality of pupils learning and their progress	.88		
The quality of learning for pupils with special educational needs and/or disabilities and their progress	.80		
The extent of pupils spiritual, moral, social and cultural development	.72		.35
The extent to which pupils adopt healthy lifestyles	.63		.31
The extent to which pupils feel safe	.76		
Pupils attendance	.56	-.49	
Pupils behaviour	.79		
The extent to which pupils contribute to the school and wider community	.73		.36
The extent to which pupils develop workplace and other skills that will contribute to their future economic well-being	.84		
The quality of teaching	.85		
The extent to which the curriculum meets pupils needs, including, where relevant, through partnerships	.81		
The effectiveness of care, guidance and support	.77		
The use of assessment to support learning	.74		
The effectiveness of leadership and management in embedding ambition and driving improvement	.88		
The effectiveness with which the school promotes equality of opportunity and tackles discrimination	.83		
The effectiveness with which the school promotes community cohesion	.62		
The effectiveness with which the school deploys resources to achieve value for money	.92		
The effectiveness of the governing body in challenging and supporting the school so that weaknesses are tackled decisively...	.70	.37	
The effectiveness of safeguarding procedures	.47	.54	
The effectiveness of the school's engagement with parents and carers	.68		
The leadership and management of teaching and learning	.82		

Supplementary Figures

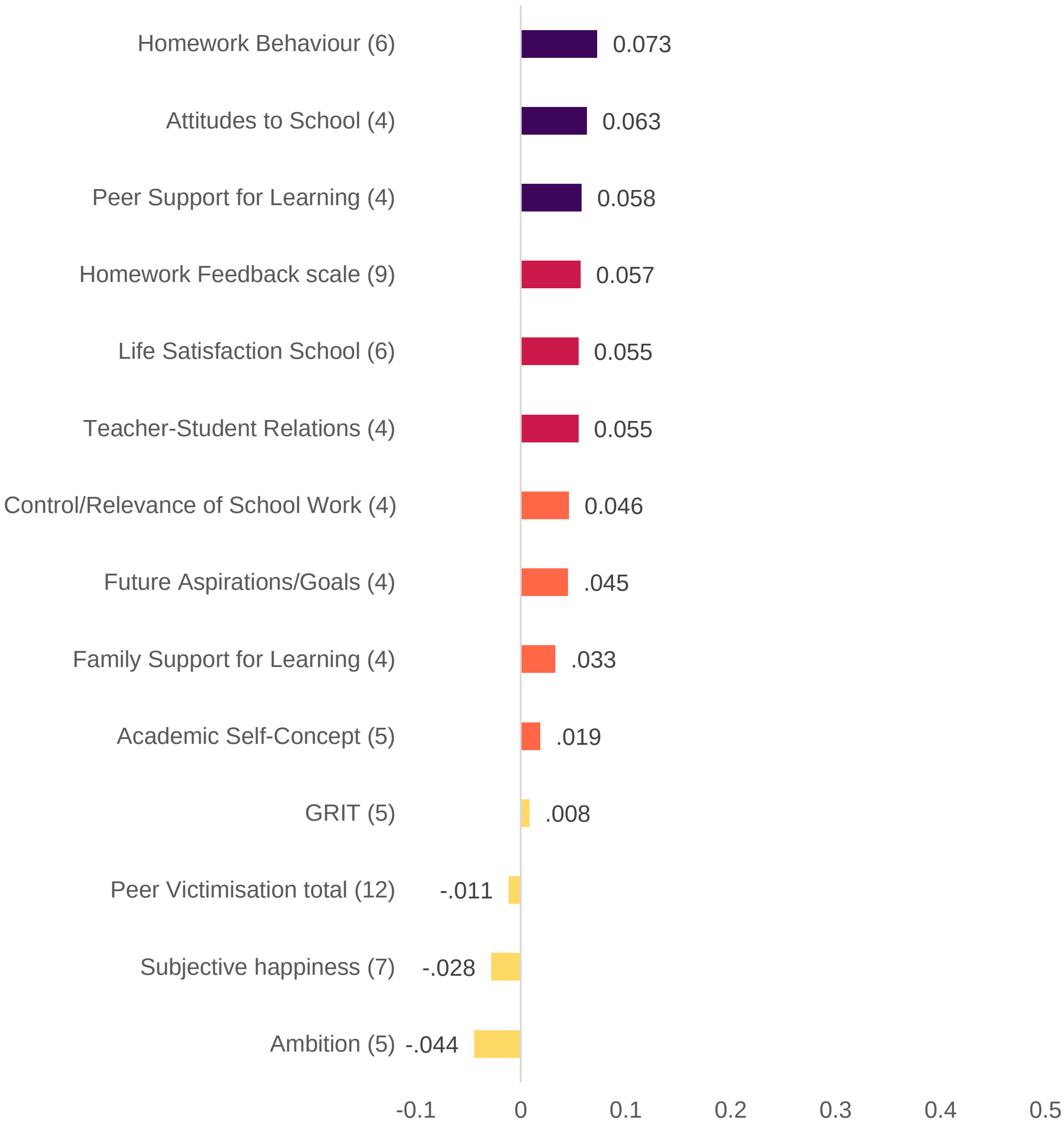


Figure S1. Spearman correlation coefficients (r_s) for the relationship between measures of student wellbeing and engagement and the Ofsted headline quality rating. The Ofsted measure was rated from 1 (inadequate) to 4 (outstanding). The total score for each of the student quality and engagement measures are in brackets. Details of how they are measured are in the Supplementary Measures section. After correcting for multiple testing (0.05 / number of correlations: 12 = .0035), only Homework behaviour was significant.

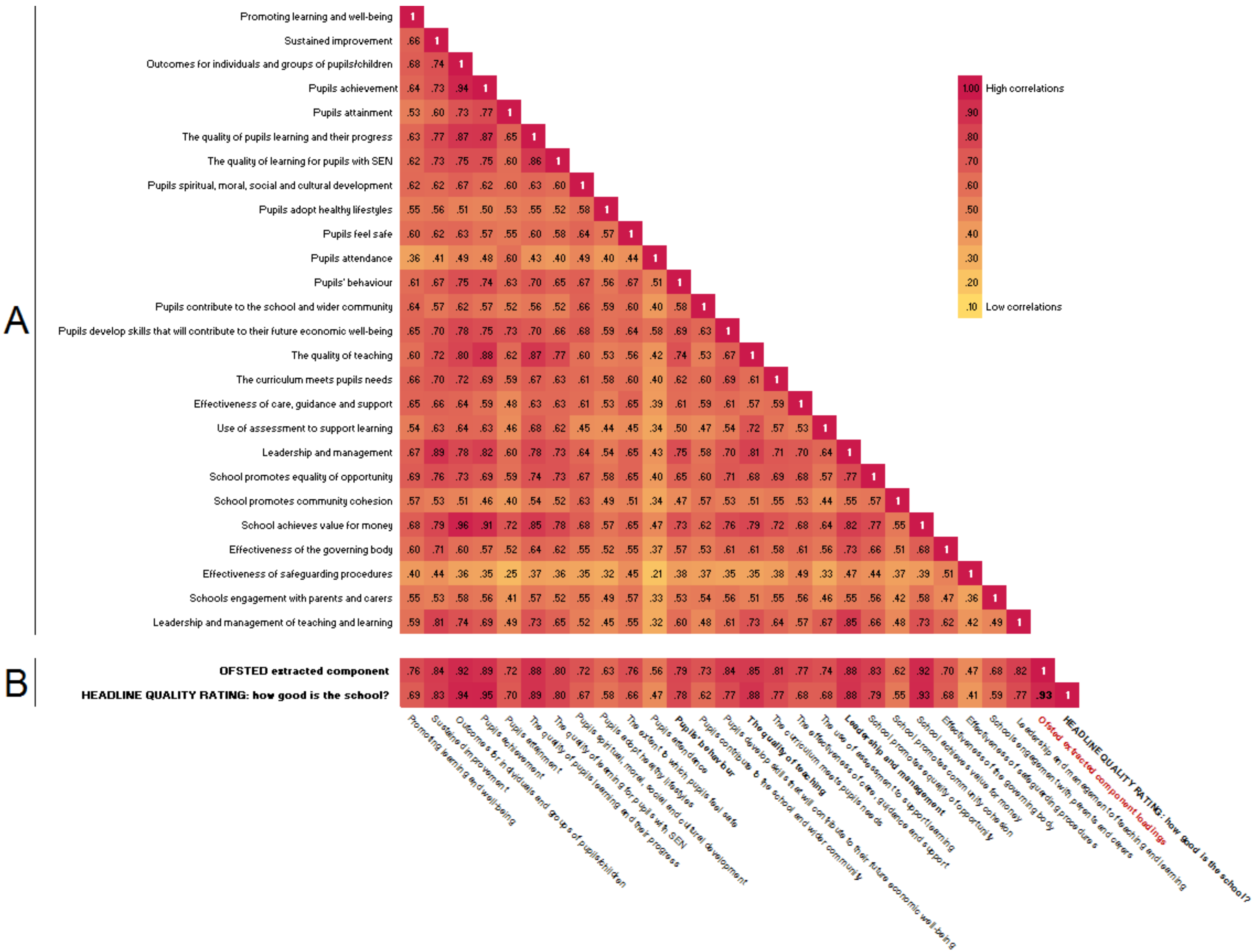


Figure S2. Correlation matrix of Ofsted individual items. **Part A** shows the intercorrelations among the 23 individual items. **Part B** shows how these individual items correlate with the Ofsted extracted component and the Ofsted Headline quality rating.

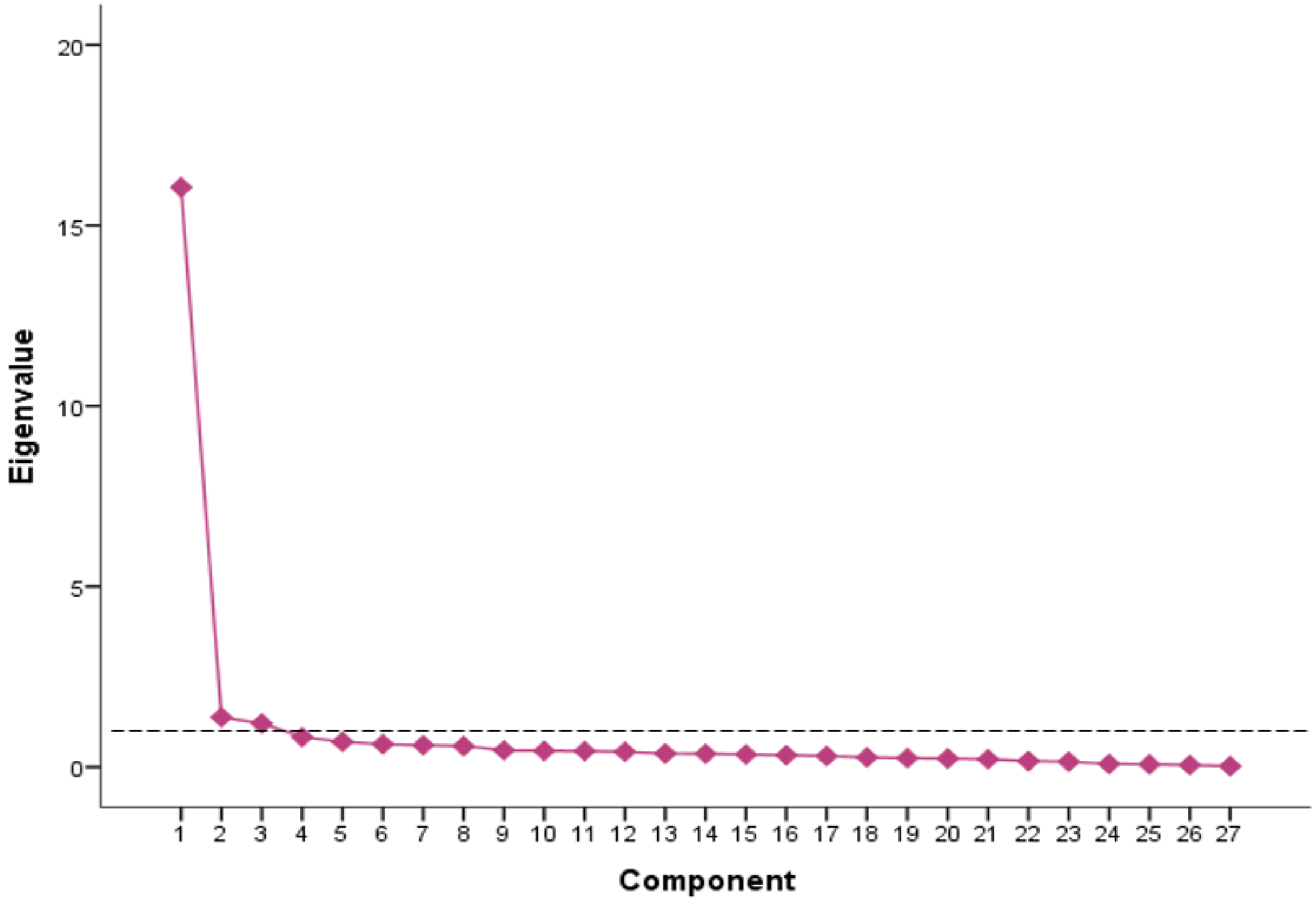


Figure S3. Scree plot showing eigenvalues for each principal component after performing PCA on individual Ofsted items. The dashed line represents the cut-off for principal component retention based on the Kaiser's λ > 1 criterion.

Appendix 4: Supplementary materials for Chapter 5

The genetics of university success

Smith-Woolley, E., Ayorech, Z., von Stumm, S., Dale, P. S., Plomin, R

Supplementary Tables

Supplementary Figures

Table S1. Mean (standard deviation) of each of the university success variables for the total sample and across gender and zygosity

			Means									ANOVA			
Uni success	***N* whole**	***N GPS***	**Whole sample**	**Male**	**Female**	**MZm**	**DZm**	**MZf**	**DZf**	**Dzos**	**DZss**	**Sex**	**Zyg**	**Sex x Zyg**	**R^2**
Entrance exam achiev.	9407	3501	322.73 (129.91)	310.96 (136.86)	332.1 (123.32)	315.04 (137.92)	308.05 (138.86)	329.15 (125.33)	334.28 (125.71)	322.41 (126.41)	322.4 (132.46)	61.89**	0.18	10.13**	<0.01
University quality	6091	2251	45.97 (32.81)	44.00 (32.11)	47.44 (33.25)	44.12 (32.03)	43.69 (31.98)	47.81 (33.25)	46.43 (32.91)	46.27 (33.07)	45.25 (32.53)	16.43**	0.44	2.79**	<0.01
University achiev.	3219	1291	4.21 (0.66)	4.22 (0.67)	4.21 (0.65)	4.27 (0.67)	4.27 (0.46)	4.23 (0.64)	4.19 (0.68)	4.21 (0.65)	4.21 (0.68)	0.06	0.14	1.05	<0.01
University enrolment	10288	3774	Y= 57%	Y=52%	Y=61%	Y=51%	Y=50%	Y=61%	Y=60%	Y=58%	Y=55%				
Reg Uni qual	5918	2179	0.00 (1.00)	-0.09 (0.95)	0.07 (1.03)	-0.06 (0.91)	-0.10 (0.96)	0.10 (1.04)	0.04 (1.03)	-0.02 (1.01)	-0.02 (1.00)	9.53*	0.95	0.20	<.001

Note: *N whole* = total sample size (both twins in a pair) after exclusions; *N GPS* = total sample size for the genotyped sample; *MZ*= monozygotic; *DZ*= dizygotic; *m*= male; *f*= female; *os*= opposite sex; Entrance exam achiev = university entrance exam grades calculated by converting achievement scores on the General Certificate of Education Advanced Level into Universities and Colleges Admissions Service (UCAS) points; University Quality= quality of university attended based on the UK university league tables in 2014; University Achiev= grade achieved at university, graded from 1 (a pass, the lowest possible pass) to 5 (a first-class degree, the highest possible pass); University enrolment = proportion of sample that either did or did not go on to university Y/N (no further analysis was done with this variable as it is dichotomous); Y= total number of individuals who went on to university in per cent; Reg uni qual. = saved standardised residuals following regression of university entrance exam achievement on university quality. Analyses of variance (ANOVA) performed on raw data from one randomly selected twin per pair to test the effect of sex, zygosity and their interaction. Results = *F* statistic; * = p<.05; ** = p<.01; R^2= proportion of variance explained by sex, zygosity and their interaction.

Table S2. Sensitivity analyses comparing entrance exam achievement and university quality between those individuals who did or did not report their final university degree grade

	Data not present Mean (*SD*)	**Data present Mean (*SD*)**	***t***	***df***	**Mean difference**	***R²***
Entrance exam achievement	0.26 (0.86)	0.46 (0.99)	-8.87**	5550	-0.21	0.01
University quality	0.03 (0.87)	-0.07 (1.02)	3.56**	5445	0.10	<0.01

Note: *t*-test comparing achievement and university quality differences between individuals who did or did not report their final university degree grade but previously stated that they were attending university. We found significant mean differences due to large sample sizes, however only 1% of variance was explained by group membership; ** p <.001.

Table S3. Twin correlations (sample size) for university success variables across gender and zygosity.

	Twin correlations								Falconer's formula		
	MZ full	**DZ full**	**MZm**	**DZm**	**MZf**	**DZf**	**Dzos**	**DZss**	**A**	**C**	**E**
Entrance exam achiev.	0.69 (N=3,339)	0.38 (N=6,068)	0.72 (N= 1,355)	0.40 (N=1,377)	0.67 (N=1,984)	0.42 (N=1,663)	0.35 (N=3,028)	0.41 (N=3,040)	0.62	0.07	0.31
Uni. enrolment*	0.66 (N=3,591)	0.42 (N=6,697)	0.70 (N= 1,576)	0.43 (N=1,579)	0.62 (N=2,015)	0.44 (N=1,785)	0.39 (N=3,333)	0.44 (N=3,364)	0.48	0.18	0.34
Uni. quality	0.65 (N=2,130)	0.34 (N=3,961)	0.69 (N= 842)	0.41 (N=835)	0.61 (N=1,288)	0.42 (N=1,103)	0.28 (N=2,023)	0.42 (N=1,938)	0.62	0.03	0.35
Uni. achievement	0.30 (N=1,222)	0.10 (N=1,997)	0.29 (N= 370)	0.04 (N=310)	0.33 (N=852)	0.09 (N=675)	0.12 (N=1,012)	0.08 (N=985)	0.30	0.00	0.70
Uni. quality regressed for prior achievement	0.50 (N=1,636)	0.26 (N=2,550)	0.51 (N=648)	0.34 (N=528)	0.48 (N=988)	0.31 (N=740)	0.15 (N=1282)	0.37 (N=1268)	0.48	0.02	0.50

Note: MZ = monozygotic; DZ = dizygotic; m= male; f= female; os= opposite sex; ss= same sex; *N*=sample size (individuals) after exclusions; * = university enrolment was a dichotomous variable (university yes or no), therefore we calculated tetrachoric correlations. It is possible to calculate rough estimates of additive genetic influence (A), shared environmental influence (C) and non-shared environmental influence (E) by using Falconer's formula. A = 2 * (rMZ - rDZ); C = rMZ – A; E = 1 – rMZ.

Table S4. Univariate model fitting results for entrance exam achievement

Model	ep	-2LL	df	AIC	∆ LL	∆ df	*p*
1. Full saturated	10	52571.27	19506	13559.27	-	-	-
2. Constrained means across twin order	8	52571.27	19508	13555.27	4.79e-09	2	1.000
3. Constrained means and variances across twin order	6	52571.27	19510	13551.27	8.73e-11	4	1.000
4. Constrained means and variances across twin order and zygosity	4	52574.85	19512	13550.85	3.75	6	0.730
ACE model	4	52574.85	19512	13550.85	3.57	6	0.73
AE model	3	52598.27	19513	13572.27	23.42	1	1.30-06
CE model	3	52976.81	19513	13950.81	401.96	1	2.06e-89

Note: ep= estimated parameters; -2LL= -2 log-likelihood; df= degrees of freedom; AIC= Akaike's information criteria; ∆LL= change in log-likelihood; ∆df= change in degrees of freedom. The ACE model fit the data best. Estimated parameters in the full model: 4 means (MZ twin 1 mean, MZ twin 2 mean, DZ twin 1 mean, DZ twin 2 mean), 4 standard deviations 'SD' (MZ twin 1 SD, MZ twin 2 SD, DZ twin 1SD, DZ twin 2 SD), 2 within twin correlations (MZ and DZ). This Saturated model is then constrained across twin means, variances, twin order and zygosity (4 estimated parameters) to test whether twin modelling assumptions have been met.

Table S5. Univariate model fitting results for university quality

Model	ep	-2LL	df	AIC	∆ LL	∆ df	*p*
1. Full saturated	10	33463.26	12248	8967.26	-	-	-
2. Constrained means across twin order	8	33463.26	12250	8963.26	-1.84e-08	2	1.000
3. Constrained means and variances across twin order	6	33463.26	12252	8955.44	-2.01e-08	4	1.000
4. Constrained means and variances across twin order and zygosity	4	33463.26	12254	8955.44	0.18	6	1.000
ACE model	4	33463.44	12254	8955.44	0.18	6	1.000
AE model	3	33468.11	12255	8958.11	4.67	1	0.031
CE model	3	33651.62	12255	9141.62	188.18	1	7.94e-43

Note: ep= estimated parameters; -2LL= -2 log-likelihood; df= degrees of freedom; AIC= Akaike's information criteria; ∆LL= change in log-likelihood; ∆df= change in degrees of freedom. The ACE model fit the data best. Estimated parameters in the full model: 4 means (MZ twin 1 mean, MZ twin 2 mean, DZ twin 1 mean, DZ twin 2 mean), 4 standard deviations 'SD' (MZ twin 1 SD, MZ twin 2 SD, DZ twin 1SD, DZ twin 2 SD), 2 within twin correlations (MZ and DZ). This Saturated model is then constrained across twin means, variances, twin order and zygosity (4 estimated parameters) to test whether twin modelling assumptions have been met.

Table S6. Univariate model fitting results for university achievement

Model	ep	-2LL	df	AIC	∆ LL	∆ df	*p*
1. Full saturated	10	18610.94	6652	5306.94	-	-	-
2. Constrained means across twin order	8	18610.94	6654	5302.94	2.31e-08	2	1.000
3. Constrained means and variances across twin order	6	18610.94	6656	5298.94	1.13e-10	4	1.000
4. Constrained means and variances across twin order and zygosity	4	18613.12	6658	5297.11	2.18	6	0.903
ACE model	4	18613.12	6658	5297.11	2.18	6	0.903
AE model	3	18613.17	6659	5295.17	-3.64e-11	1	1.000
CE model	3	18652.13	6659	5334.13	0.39	1	4.32e-10

Note: ep= estimated parameters; -2LL= -2 log-likelihood; df= degrees of freedom; AIC= Akaike's information criteria; ∆LL= change in log-likelihood; ∆df= change in degrees of freedom. The AE model fit the data best. Estimated parameters in the full model: 4 means (MZ twin 1 mean, MZ twin 2 mean, DZ twin 1 mean, DZ twin 2 mean), 4 standard deviations 'SD' (MZ twin 1 SD, MZ twin 2 SD, DZ twin 1SD, DZ twin 2 SD), 2 within twin correlations (MZ and DZ). This Saturated model is then constrained across twin means, variances, twin order and zygosity (4 estimated parameters) to test whether twin modelling assumptions have been met.

Table S7. Univariate model fitting results for university quality independent of entrance exam achievement

Model	ep	-2LL	df	AIC	∆ LL	∆ df	*p*
1. Full saturated	10	32893.71	11798	9297.71	-	-	-
2. Constrained means across twin order	8	32893.71	11800	9293.71	5.44e-08	2	1.000
3. Constrained means and variances across twin order	6	32893.71	11802	9289.7131	-2.62e-10	4	1.000
4. Constrained means and variances across twin order and zygosity	4	32898.52	11804	9290.52	4.81	6	0.569
ACE model	4	32898.52	11804	9290.52	4.81	6	0.569
AE model	3	32898.93	11805	9288.93	0.41	1	0.520
CE model	3	32979.92	11805	9369.92	81.41	1	1.84e-19

Note: ep= estimated parameters; -2LL= -2 log-likelihood; df= degrees of freedom; AIC= Akaike's information criteria; ∆LL= change in log-likelihood; ∆df= change in degrees of freedom. The AE model fit the data best. Estimated parameters in the full model: 4 means (MZ twin 1 mean, MZ twin 2 mean, DZ twin 1 mean, DZ twin 2 mean), 4 standard deviations 'SD' (MZ twin 1 SD, MZ twin 2 SD, DZ twin 1SD, DZ twin 2 SD), 2 within twin correlations (MZ and DZ). This Saturated model is then constrained across twin means, variances, twin order and zygosity (4 estimated parameters) to test whether twin modelling assumptions have been met.

Table S8. Liability threshold model fitting results for university enrolment

Model	ep	-2LL	df	AIC	Δ LL	Δ df	*p*
Saturated model	6	26712.83	21304	-15895.17	-	-	-
Comparison with Sub 1 model	4	26712.83	21306	-15899.17	2.10×10^{-4}	2	0.99
Comparison with Sub 2 model	3	26712.93	21307	-15901.07	9.96×10^{-2}	3	0.99

Note: Sub 1= constrained model equating thresholds across Twin 1 and Twin 2 within zygosity groups; Sub 2 = constrained model equating Thresholds across Twin 1 and Twin 2 and zygosity group; ep= estimated parameters; -2LL= -2 log-likelihood; df= degrees of freedom; AIC= Akaike's information criteria; ΔLL= change in log-likelihood; Δdf= change in degrees of freedom.

Table S9. Twin *ACE* estimates and 95% confidence intervals indicating the proportion of variance in university success variables attributable to genetic and environmental factors.

	A	**C**	**E**
Entrance exam achievement	0.57 (0.52-0.63)	0.12 (0.07-0.17)	0.31 (0.30-0.33)
University enrolment	0.51 (0.44-0.58)	0.36 (0.29-0.42)	0.13 (0.11-0.15)
University quality (total score)	0.57 (0.49-0.65)	0.08 (0.01-0.14)	0.35 (0.32-0.37)
University achievement	0.46 (0.33-0.52)	0.00 (0.00-0.11)	0.53 (0.49-0.58)
University quality regressed for entrance exam achievement	0.47 (0.37-0.54)	0.03 (0.00-0.10)	0.50 (0.47-0.54)

Note: A= Additive genetic influence, C= Common (shared) environmental influence, E= Non-shared environmental influence. If confidence intervals include 0 then the estimate was not significant.

Table S10. Multivariate model fitting results for university entrance exam achievement, university quality and university achievement. Comparing correlated factors, independent pathway and common pathway models.

Model	ep	-2LL	df	AIC	∆ LL	∆ df	*p*
Full saturated	54	100856.0	38700	23455.98	-	-	-
Correlated factors	21	100872.7	38733	23406.75	16.76	33	0.99
Independent pathway	21	100880.5	38733	23414.50	24.52	33	0.86
Common pathway	18	100921.1	38737	23447.13	65.15	37	0.002

Note: ep= estimated parameters; -2LL= -2 log-likelihood; df= degrees of freedom; AIC= Akaike's information criteria; ∆LL= change in log-likelihood; ∆df= change in degrees of freedom. The correlated factors solution fit the data best.

Table S11. Genetic, shared environmental and non-shared environmental correlations between university success measures

a. Genetic correlations between university success variables

	Entrance_achiev.	Uni_Quality	Uni_Achiev.
Entrance_achiev.	1.00		
Uni_Quality	0.76	1.00	
Uni_Achiev.	0.49	0.27	1.00

b. Shared environmental correlations between university success variables

	Entrance_achiev.	Uni_Quality	Uni_Achiev.
Entrance_achiev.	1.00		
Uni_Quality	0.81	1.00	
Uni_Achiev.	0.35	0.27	1.00

c. Non-shared environmental correlations between university success variables

	Entrance_achiev.	Uni_Quality	Uni_Achiev.
Entrance_achiev.	1.00		
Uni_Quality	0.35	1.00	
Uni_Achiev.	0.03	0.09	1.00

Table S12. Polygenic score prediction (R^2) of university success variables with 95% confidence intervals

	R^2	95% CIs	
Entrance exam achievement	.041	.028	.054
University attainment	.053[a]	.039	.067
University quality	.023	.010	.035
University achievement	.007	-.002	.016

Note: CIs = confidence intervals; [a]= Variable was dichotomous, therefore Nagelkerke R^2 was calculated.

Table S13. Sample sizes and EduYears GPS prediction (R^2) for STEM and humanities subjects

	N whole	*N* GPS	STEM	Humanities
natural sciences	517	210	√	
mathematics/statistics	124	44	√	
medicine/veterinary	133	45	√	
engineering	139	50	√	
technology/design	98	50	√	
computing/IT	88	30	√	
social sciences	498	194		√
arts	429	162		√
humanities	404	167		√
languages	96	36		√
law	109	36		√
Total (whole sample)	2635		1099	1536
Total (GPS)		1024	429	595
GPS prediction of degree achievement (R^2)			0.003	0.015

Note: N whole = total sample size (individuals) with exclusions; N GPS = sample size for GPS analysis; STEM = Science, technology, engineering and mathematics. We compared the *EduYears* GPS prediction of STEM degree achievement to Humanities degree achievement using Fisher's r-to-z transformation. We found no significant difference between the predictions (z = 1.08, p = 0.28 two tailed).

Supplementary Figures

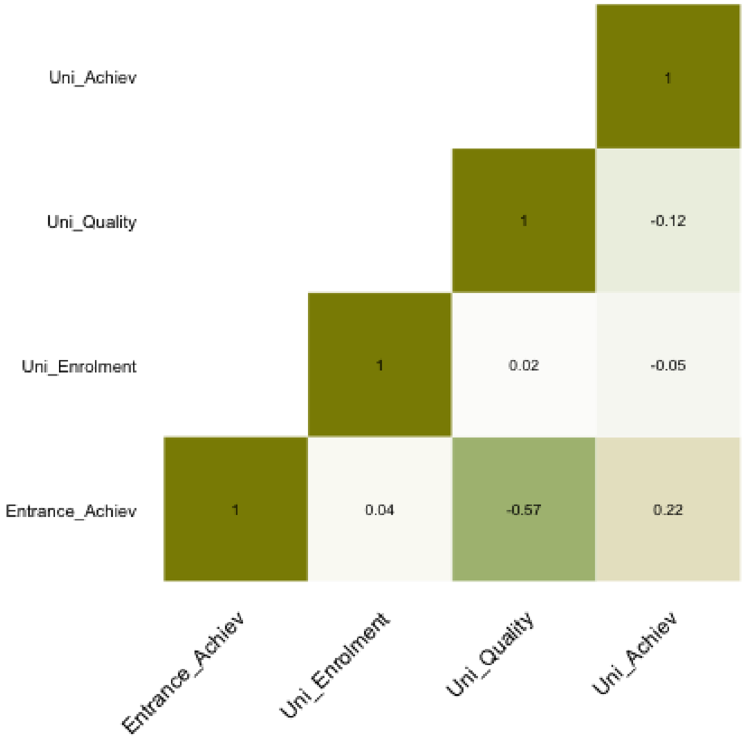


Figure S1. Correlations between each of the university success variables. ***Note:*** Entrance_Achiev= entrance exam achievement; Uni_Enrolment= university enrolment; Uni_Quality= university quality and Uni_Achiev= university achievement.

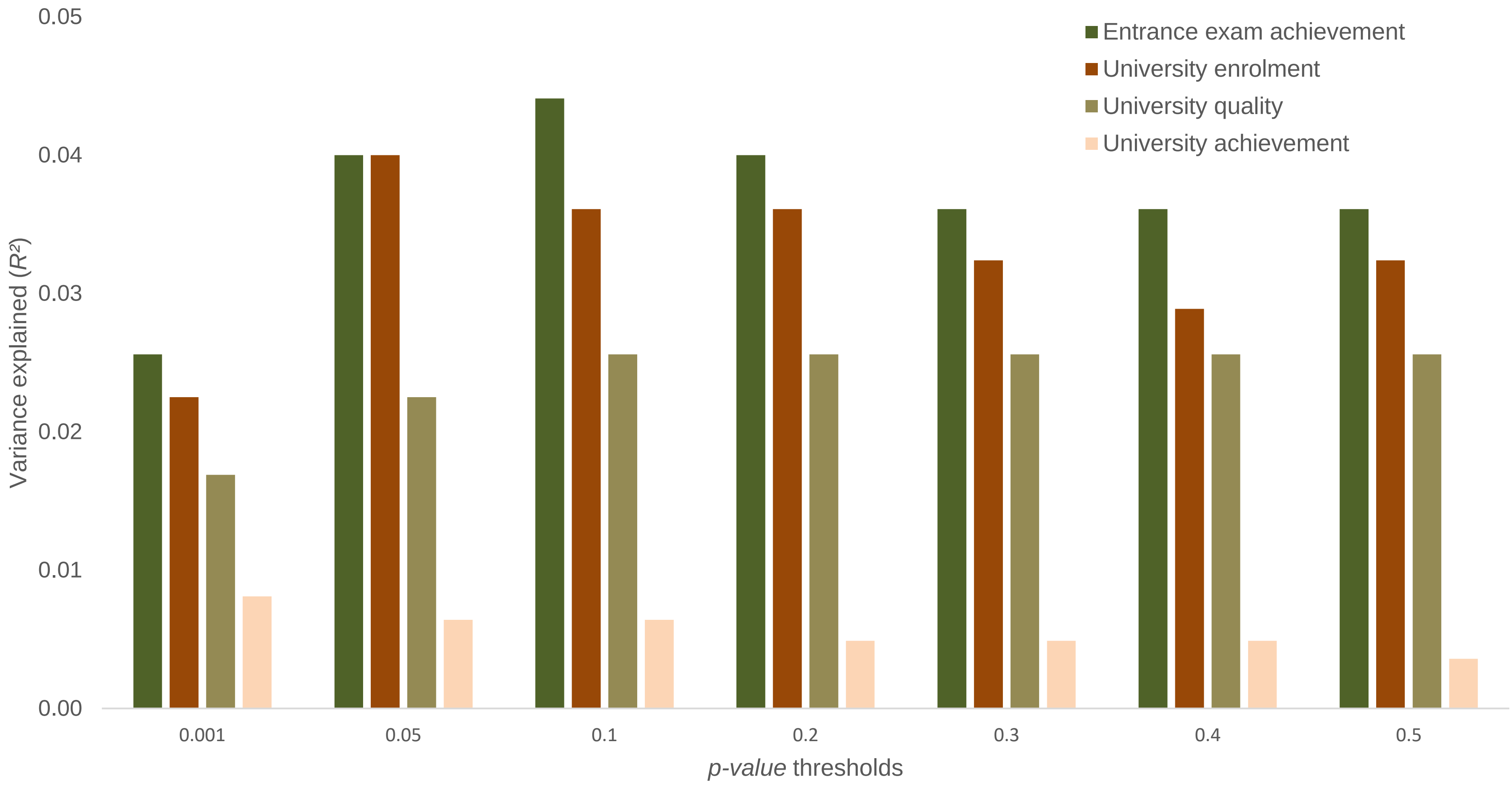


Figure S2. Polygenic score prediction for each of the university success variables at each of *p-value* thresholds tested using PRSice.